Identification of markers for dietary intake and health status by GC-MS based metabolite profiling approaches

Dissertation zur Erlangung des Doktorgrades der Naturwissenschaften (Dr. rer. nat.)

Fakultät Naturwissenschaften
Universität Hohenheim

Max Rubner-Institut, Institut für Sicherheit und Qualität bei Obst und Gemüse

Universität Hohenheim, Institut für Lebensmittelchemie

vorgelegt von

Carina I. Mack

aus *Ostfildern Ruit*

2019

Dekan:	Prof. Dr. Uwe Beifuß
1. berichtende Person:	Prof. Dr. Sabine E. Kulling
2. berichtende Person:	Prof. Dr. Walter Vetter
Eingereicht am:	22.03.2019
Mündliche Prüfung am:	19.06.2019

TABLE OF CONTENTS

ZUSAMMENFASSUNG

Marker sind Substanzen, die als Indikatoren für eine Exposition, einen metabolischen Zustand oder einen Effekt herangezogen werden. Metabolomics und *Metabolite profiling*-Ansätze gewinnen in der Markerforschung zunehmend an Bedeutung. Metabolomics ermöglicht die Identifizierung von Markern für den Lebensmittelverzehr, die in Zukunft unter anderem in epidemiologischen Studien zur Ergänzung und Überprüfung traditioneller Ernährungserhebungsmethoden Verwendung finden werden. In der Konsequenz können Zusammenhänge zwischen Ernährung und Gesundheit im Allgemeinen, sowie ernährungsabhängigen Erkrankungen im Speziellen, besser beschrieben werden. Außerdem können mittels Metabolomics auch Marker identifiziert werden, die im klinischen Rahmen eine Diagnose, Prognose oder Überwachung von Behandlungsmaßnahmen für eine Erkrankung ermöglichen, wie z.B. bei Typ 2 Diabetes mellitus. Von besonderem Interesse sind dabei Marker, die eine frühzeitige Diagnose, das heißt vor der Manifestation von Symptomen, ermöglichen. Ziel der vorliegenden Arbeit war es, die Verwendung von GC-MS basierten *Metabolite profiling*-Ansätzen zur Identifizierung von Markern für den Lebensmittelverzehr und den Gesundheitsstatus zu prüfen. Einen besonderen Schwerpunkt bildeten dabei volatile organische Verbindungen und Zuckerverbindungen, die in Urin- und Plasmaproben einer Querschnittsstudie mit 300 Probanden sowie einer humanen Interventionsstudie mit Diabetikern, Prädiabetikern und Gesunden analysiert wurden.

Bei der bisherigen Suche nach Markern für den Lebensmittelverzehr lag das Augenmerk vor allem auf nicht-volatilen Metaboliten. Um das Potential des Volatiloms zu eruieren, wurden Urinproben einer Querschnittsstudie mithilfe einer ungerichteten HS-SPME-GC×GC-MS Methode analysiert und darin beispielhaft nach Markern für den Kaffeekonsum gesucht. Aus dem Urinprofil mit 138 volatilen Verbindungen wurden sechs plausible Kandidaten identifiziert, von denen sich 3,4-Dimethyl-2,5-furandion als der robusteste Marker erwies. Mittels einer Korrelationsanalyse anhand von Verzehrsdaten weiterer Lebensmittel wurde darüber hinaus gezeigt, dass das Volatilom eine vielversprechende Quelle neuer Marker für den Lebensmittelverzehr ist.

Zucker lassen sich aufgrund ihrer strukturellen Ähnlichkeit häufig nur unzureichend chromatographisch trennen, daher werden in humanen Matrices bisher mehrheitlich nur wenige bekannte Zuckerverbindungen erfasst. Im Rahmen dieser Arbeit wurde eine semi-gerichtete GC-MS Zuckerprofiling-Methode entwickelt, mit der gezeigt werden konnte, dass das humane Zuckerprofil im Urin und im Plasma erheblich kom-

plexer ist, als bisher beschrieben und angenommen. Verschiedene Zuckerverbindungen, wie beispielsweise Psicose oder Trehalose, über deren Herkunft und Vorhandensein im Urin oder in Plasma fast nichts bekannt ist, wurden nachgewiesen. Im Urin zeigten sich darüber hinaus Unterschiede in der Maltosekonzentration in Abhängigkeit vom Geschlecht sowie dem prä- und postmenopausalen Status, die vermutlich im Zusammenhang mit der vaginalen Mikrobiota stehen. Die Assoziation der Zuckerprofile mit dem Lebensmittelverzehr ermöglichte zudem die Identifizierung neuer und Bestätigung bekannter Marker, beispielsweise für den Verzehr von Avocado und Milchprodukten, sowie für Alkoholkonsum.

Im Plasma von Gesunden, Prädiabetikern und Diabetikern wurden nach einem oralen Glucosetoleranztest deutliche Unterschiede im Zuckerprofil festgestellt. Interessanterweise zeigten eine Reihe neuer Zuckerverbindungen markante postprandiale Unterschiede abhängig vom Gesundheitszustand. Beispielsweise zeigte Trehalose ein ähnliches Profil wie die insulinabhängige Glucose. Jedoch ist weder über den Mechanismus noch zur Herkunft dieser Zucker etwas bekannt.

Bereits die bisherigen Ergebnisse des Zuckerprofilings in Urin und Plasma zeigten, dass zusätzliche Zuckerverbindungen, wenn auch in sehr geringer Konzentration, vorhanden sind. Daher wurde die eindimensionale Methode zu einer zweidimensionalen GC×GC-MS-Methode mit verbesserter Sensitivität und Trennung weiterentwickelt, was nun die Erfassung von 84 statt 55 Zuckerverbindungen in Urin ermöglicht. Erste Auswertungen der Messdaten einer Interventionsstudie mit Äpfeln zeigten, dass diese Methode die Identifizierung von potentiellen Markern für den Verzehr von Äpfeln ermöglicht. Die Ergebnisse verdeutlichen, welches Potential in der umfassenden Analyse von Zuckern, einschließlich seltener Verbindungen, steckt.

GC-MS basierte *Metabolite profiling*-Ansätze, wie hier für das Volatilom und das Zuckerprofil gezeigt, sind geeignete Methoden für die Identifizierung von Markern des Lebensmittelverzehrs und des Gesundheitsstatus. Die Identifizierung bisher unbekannter Verbindungen, die Weiterentwicklung der Zuckeranalytik zu einer quantitativen Methode und insbesondere die Validierung der identifizierten Marker bezüglich ihrer Eignung, den Lebensmittelverzehr bzw. den diabetischen Status akkurater zu erfassen, sind zukünftige Ziele. Besonders herausfordernd ist es dabei, die mechanistischen Zusammenhänge aufzuklären, insbesondere im Hinblick auf Herkunft, Vorhandensein und Funktion der detektierten Zuckerverbindungen im menschlichen Metabolismus.

SUMMARY

Markers are compounds that can be used as indicators of an exposure, a metabolic state, or any other effect. Metabolomics and metabolite profiling approaches for marker discovery will increasingly gain significance. In the context of food, diet, and health, these approaches allow among others the identification of dietary intake markers, which can complement and verify traditional dietary assessment methods in epidemiologic studies. Consequently, the investigation of associations between diet and health status in general, and also in particular diet-related diseases will be improved. Compared to classical biomarker studies, metabolomics enables a more comprehensive investigation of clinical markers for diagnosis, prognosis and monitoring of diseases, such as type 2 diabetes mellitus. Especially, early diagnosis in pre-disease states, where symptoms are not yet evident, are of particular interest. The aim of this thesis was to evaluate the application of GC-MS based metabolite profiling approaches for the identification of markers for dietary intake and health status. In this respect, volatile organic compounds and sugar compounds were analyzed to discover marker candidates in urine and plasma samples from a cross-sectional study with 300 participants, as well as from a human intervention study with diabetic, prediabetic and healthy participants.

In the past, the search for markers of dietary intake mostly focused on non-volatile metabolites. To explore the potential of the volatilome, urine samples of a cross-sectional study were analyzed aiming to exemplary identify markers of coffee consumption using an untargeted HS-SPME-GC×GC-MS method. Six marker candidates were identified from a profile of 138 volatile organic compounds with the most robust represented by 3,4-dimethyl-2,5-furandione. Moreover, the correlation with the general dietary intake data highlighted the volatilome as a particularly interesting source for the detection of new dietary markers.

The chromatographic separation of sugar compounds is often insufficient due to the high structural similarities. Therefore, in most studies common and well-known sugar compounds are analyzed in human body fluids. Within the scope of this thesis, a semitargeted GC-MS sugar profiling method was developed, which revealed a more complex sugar profile, both in urine and plasma, than described so far or expected. Rare sugar compounds such as psicose and trehalose were detected. However, the knowledge about their origin and presence in urine or plasma is limited to date. Moreover, the maltose concentration in urine was shown to be dependent on sex and

menopause status (pre- and post-menopausal) – a relationship with the vaginal microbiota is suggested here. In addition, the association of the urinary sugar profile with dietary intake data enabled the identification and confirmation of several new and also known marker candidates as for example, for consumption of avocado, dairy products and alcohol.

The plasma sugar profiles of healthy, prediabetic and diabetic volunteers after an oral glucose tolerance test could be clearly distinguished, independent of glucose. Remarkably, a variety of sugar compounds showed marked postprandial differences dependent on health status. For example, trehalose showed a profile similar to the insulin-dependent profile of glucose. However, the origin and underlying biological mechanism for those sugar compounds remain to be elucidated.

During the application of the one-dimensional GC-MS sugar profiling method to urine and plasma samples, it became evident that even more sugar compounds might be present, although in low concentrations, but were not detected due to limitations of the analytical method. Therefore, the one-dimensional method was transferred into a two-dimensional GC×GC-MS method. Improved sensitivity and separation finally enabled the detection of 84 instead of 55 sugar compounds in urine. The two-dimensional method was applied in an intervention study with apples, and revealed marker candidates for apple consumption for future validation. Overall, the results illustrate the benefit of a comprehensive analysis of sugar compounds in urine and plasma, including minor and rare sugar derivatives.

The GC-MS based metabolite profiling approaches addressing the volatilome and the sugar profile, respectively, were demonstrated to be promising approaches for the identification of markers for dietary intake and health status. Future work should address the identification of unknown compounds, the adaptation of the GC×GC-MS sugar profiling method for quantitative purposes, and especially the validation of the identified marker candidates with respect to their suitability to more accurately assess dietary intake or diabetic state. High priority should also be given to the biochemical mechanisms and the origin of the compounds as well as their physiological or pathophysiological function in human metabolism.

ABBREVIATIONS

ANOVA	analysis of variance
APCI	atmospheric pressure chemical ionization
BSA	benzenesulfonic acid
BSTFA	*N,O*-bis(tri-methylsilyl)trifluoroacetamide
CAR	carboxen
CE-MS	capillary electrophoresis coupled to mass spectrometry
CI	chemical ionization
CV	coefficient of variation
CW	carbowax
C18	octadecylsilyl derivatized silica column
^{1}D	first dimension
^{2}D	second dimension
2D	two-dimensional
d_c	internal diameter
d_f	film thickness
DVB	divinylbenzene
ECD	electron capture detector
EI	electron ionization
FAME	fatty acid methyl ester
FFQ	food-frequency questionnaire
FID	flame ionization detector
FoodBAll	Food Biomarker Alliance
GC	gas chromatography
GC×GC	comprehensive two-dimensional gas chromatography
GMD	Golm Metabolome Database

HbA_{1C}	glycated hemoglobin
HDL	high-density lipoprotein
HLB	hydrophilic lipophilic balances
HMDB	Human Metabolome Database
HPLC	high pressure liquid chromatography
HS	headspace
IFG	impaired fasting glycaemia
IGT	impaired glucose tolerance
IS	internal standard
L	length
LC	liquid chromatography
LDL	low-density lipoprotein
LLE	liquid-liquid extraction
MRI	Max Rubner-Institut
MSI	Metabolomics Standard Initiative
MS/qMS	mass spectrometry/quadrupole mass spectrometer
MSTFA	*N*-methyl-*N*-(rimethylsilyl)trifluoroacetamide
m/z	mass-to-charge ratio
NIST	National Institute of Standards and Technology
NMR	nuclear magnetic resonance spectroscopy
OGTT	oral glucose tolerance test
PA	polyacryl
PAN	polyacrylonitrile
PCA	principal component analysis
PDMS	polydimethylsiloxane
PLS	partial least squares

QC	quality control
RI	retention index
RT	retention time
SIM	selected ion monitoring
SPE	solid-phase extraction
SPME	solid-phase micro extraction
S/N	signal-to-noise
TIC	total ion current
TLC	thin layer chromatography
TMCS	trimethylchlorosilane
TMS	trimethylsilyl
TOF	time-of-flight
TPR	templated resin
T1DM	type 1 diabetes mellitus
T2DM	type 2 diabetes mellitus
VOC	volatile organic compounds

PRELIMINARY REMARKS AND CONTRIBUTIONS

This doctoral thesis was carried out under the supervision of Dr. Christoph Weinert and Prof. Dr. Sabine Kulling at the Max Rubner-Institut, Karlsruhe, Germany, between February 2015 and March 2019.

Full publications as part of the doctoral thesis

1) **Mack CI**, Egert B, Liberto E, Weinert CH, Bub A, Hoffmann I, Bicchi C, Kulling SE, Cordero C. Robust markers of coffee consumption identified amongst the volatile organic compounds in human urine. Molecular Nutrition and Food Research 2019, 63, 1801060.
2) **Mack CI**, Weinert CH, Egert B, Ferrario PG, Bub A, Hoffmann I, Watzl B, Daniel H, Kulling SE. The complex human urinary sugar profile: determinants revealed in the cross-sectional KarMeN study. American Journal of Clinical Nutrition 2018; 108(3): 502-516.
3) **Mack CI**, Ferrario PG, Weinert CH, Egert B, Hoefle AS, Lee YM, Skurk T, Kulling SE, Daniel H. Exploring the diversity of sugar compounds in healthy, prediabetic and diabetic volunteers. Molecular Nutrition and Food Research 2020, 64, 1901190.

Contributions

The contributions of the authors to the full publications are as follows:

Carina I. Mack performed all essential practical sample analysis, statistical analysis and interpretation. She prepared the original manuscripts leading to the publications.

Christoph H. Weinert supervised and advised in analytical issues at the Max Rubner-Institut (Publications 2-3).

Erica Liberto and **Chiara Cordero** supervised and advised in analytical issues at the University of Turin (Publication 1).

Björn Egert was responsible for the bioinformatical processing of the obtained data.

Christoph H. Weinert, **Björn Egert** and **Paola G Ferrario** advised in issues for the statistical analysis of the obtained data.

Chiara Cordero advised with respect to the data processing of two-dimensional data using GC Image®.

Achim Bub, **Bernhard Watzl**, **Ingrid Hoffmann** and **Sabine E. Kulling** were responsible for the development and conduction of the KarMeN study. Achim Bub was responsible for the participant recruitment, characterization and sample collection. Ingrid Hoffmann was responsible for the dietary assessment during the KarMeN study.

Hannelore Daniel, **Thomas Skurk**, **Yu-Mi Lee** and **Anja S. Höfle** were responsible for the development and conduction of the MIPROMET study at the Technical University of Munich and provided the samples for publication 3.

Christoph H. Weinert, Sabine E. Kulling and **Hannelore Daniel** developed the idea behind the sugar profiling method (publications 2-3).

Sabine E. Kulling, **Christoph H. Weinert**, **Chiara Cordero** and **Carlo Bicchi** initiated the idea for the VOC analysis of human urine after coffee consumption (publication 1).

Christoph H. Weinert, Sabine E. Kulling and **Hannelore Daniel** critically reviewed the original manuscripts leading to the publications.

Sabine E. Kulling was the supervisor of this doctoral thesis.

Further publications

1) **Mack CI**, Wefers D, Schuster P, Weinert CH, Egert B, Bliedung S, Trierweiler B, Muhle-Goll C, Bunzel M, Luy B, Kulling SE. Untargeted multi-platform analysis of the metabolome and the non-starch polysaccharides of kiwifruit during postharvest ripening. Postharvest biology and technology 2017, 125: 65-76.
   - **Carina I. Mack** performed the generation of kiwifruit samples, the analysis of kiwifruit with the untargeted GC×GC-MS and sugar profiling method and prepared the original manuscript leading to this publication.
2) Rist MJ, Roth A, Frommherz L, Weinert CH, Krüger R, Merz B, Bunzel D, **Mack CI**, Egert B, *et al*. Metabolite patterns predicting sex and age in participants of the Karlsruhe Metabolomics and Nutrition (KarMeN) study. PLoS one 2017, 12: e0183228.
   - **Carina I. Mack** performed the analytical work and data processing of the sugar profiling in human urine samples of the KarMeN study.
3) Armbruster M, Rist MJ, Seifert S, Frommherz L, Weinert CH, **Mack CI**, Roth A, Merz B, Bunzel D, *et al*. European Journal of Nutrition 2018, doi: 10.1007/s00394-018-1767-1.
   - **Carina I. Mack** performed the analytical work and data processing of the sugar profiling in human urine samples of the KarMeN study.
4) Rothwell JA, Madrid-Gambin F, Garcia-Aloy M, Andres-Lacueva C, Logue C, Gallagher AM, **Mack C**, Kulling SE, Gao Q, Praticò G, *et al*. Biomarkers of intake for coffee, tea, and sweetened beverages. Genes and Nutrition 2018; 13(1): 15.
   - **Carina I. Mack** prepared the part about “sugar-sweetened beverage biomarkers” of the original manuscript leading to this publication.

Oral presentations presented by Carina Mack

1) **Mack CI**. Metabolische Veränderungen von Kiwifrüchten im Reifungsverlauf. Max Rubner-Institut-Kolloquium, February 10, 2015, Karlsruhe, Germany.
2) **Mack CI**, Weinert CH, Egert B, Hummel E, Ferrario PG, Bub A, Watzl B, Daniel H, Kulling SE. Sugar profiling analysis in nutrition. Max Rubner Conference 2016, October 10-12, 2016, Karlsruhe, Germany.

3) **Mack CI**, Weinert CH, Egert B, Hummel E, Ferrario PG, Bub A, Watzl B, Daniel H, Kulling SE. Zuckerverbindungen in Urin und Plasma: Komplexität und Einflussfaktoren. 54. Wissenschaftlicher Kongress der DGE, March 1-3, 2017, Christian-Albrechts-Universität zu Kiel, Germany.
4) **Mack CI**, Weinert CH, Egert B, Wittig F, Ferrario PG, Bub A, Liberto E, Cordero C, Watzl B, Kulling SE. Metabolitenprofile von Zuckern und Volatilen in Urin. Max Rubner-Institut-Kolloquium, April 11, 2017, Karlsruhe, Germany.
5) **Mack CI**. Oral presentation about the project status of the FoodBAll project. Annual FoodBAll Meeting, August 31-September 1, 2017, Varna, Bulgaria.
6) **Mack CI**. Oral presentation about the project status of the FoodBAll project. Annual FoodBAll Meeting, January 24-26, 2018, Berne, Swiss.
7) **Mack CI**, Weinert CH, Egert B, Daniel H, Kulling SE. Zuckerprofile in humanem Plasma von Gesunden, Prädiabetikern und Diabetikern in einem OGT-Test. 55. Wissenschaftlicher Kongress der DGE, March 7-9, 2018, Universität Hohenheim, Germany.

Poster presentations

1) **Mack CI**, Weinert CH, Egert B, Trierweiler B, Kulling SE. Metabolische Veränderungen während der Reifung von Kiwifrüchten (*Actinidia deliciosa* cv. Hayward). 44. Deutscher Lebensmittelchemikertag. September 14-16, 2015, Karlsruhe, Germany.
2) **Mack CI**, Weinert CH, Egert B, Bub A, Watzl B, Kulling SE. Semi-targeted metabolomics for the analysis of urinary sugar species: Application on the KarMeN cohort. METABO Meeting, December 7-9, 2015, Cambridge, United Kingdom.
3) **Mack CI**, Wefers D, Schuster P, Weinert CH, Egert B, Bliedung S, Trierweiler B, Muhle-Goll C, Bunzel M, Luy B, Kulling SE. Metabolische Veränderungen während der Reifung von Kiwifrüchten (*Actinidia deliciosa* cv. Hayward). 50. Vortragstagung der Deutschen Gesellschaft für Qualitätsforschung (Pflanzliche Nahrungsmittel) e.V., March 14-15, 2016, Berlin, Germany.
4) **Mack CI**, Weinert CH, Egert B, Hummel E, Bub A, Watzl B, Kulling SE. Semi-targeted GC-MS method for metabolomics analysis of urinary sugar species.

12th Annual Conference of the Metabolomics Society. June 27-30, 2016, Dublin, Ireland.

5) **Mack CI**, Egert BE, Liberto E, Weinert CH, Kulling SE, Cordero C. The urinary volatilome after coffee consumption. 5th International Conference on Foodomics, January 10-12, 2018, Cesena, Italy.
   Awarded with "Exacta-Optech" Best Poster Award for a Young Researcher.
6) **Mack CI**. Candidate dietary markers for apple and coca cola consumption. FoodBAll Finals Meeting, May 14-15, 2018, Wageningen, Netherlands.

Supervised master and diploma theses

1) Mona Ehlers. Vergleich der Metabolitenprofile von herkömmlichen und neugezüchteten Apfelsorten mithilfe ungerichteter, GC-basierter Metabolomanalysen. 2015, am Max Rubner-Institut/an der Westfälischen Wilhelms-Universität Münster.
2) Viola Kreß. Qualitatives Screening des Metaboloms verschiedener Obstsorten mittels umfassender zweidimensionaler Gaschromatographie (GC×GC-MS). 2016, am Max Rubner-Institut/an der Universität Hohenheim.
3) Julia Klein. Qualitatives Screening des Metaboloms verschiedener Gemüsearten mittels umfassender zweidimensionaler Gaschromatographie (GC×GC-MS). 2016, am Max Rubner-Institut/an der Universität Hohenheim.
4) Stefanie Lotzien. Einfluss einer kurzzeitigen fructose- bzw. glucosereichen Ernährung auf das humane Serum-Metabolom und insbesondere das Zuckerprofil. 2018, am Max Rubner-Institut/an der Rheinischen Friedrich-Wilhelms-Universität Bonn.

1 INTRODUCTION

1.1 Metabolomics

1.1.1 Principle and definitions

Metabolomics is a comprehensive analytical approach aiming at an unbiased analysis of the entirety of metabolites occurring in a biological system at a certain time point, namely the metabolome [1]. In essence, a biological system is everything that lives or lived or a sample thereof. Metabolites are organic, low molecular weight molecules (<1500 Da) in a biological system, which originate from endogenous metabolism, host-associated microbial communities, exogenous sources (e.g. environment) and are detectable [2, 3]. Metabolomics can be applied to analyze the metabolic state of a biological system (human, animal, plant, microorganisms etc.) comprehensively. In this thesis, only human biofluids were analyzed. Existing definitions of metabolomics are not consistent; the following are, however, commonly used. In metabolomics, several approaches can be distinguished with fluid transitions between them:

- **Targeted analysis** aims at a precise quantitative analysis of selected known metabolites and is in most cases driven by a specific biochemical question (hypothesis-driven). Accordingly, specific sample preparation and purification adjusted for the respective applications as well as highly sensitive and robust methods are necessary [1, 4].
- **(Targeted) Metabolite profiling** aims at the absolute or relative quantitative analysis of metabolites (known or unknown) from a selected biochemical pathway or a specific substance class. The approach can be hypothesis-driven or hypothesis-generating. Sample treatment can be specific for a substance class [1, 2, 4].
- **Untargeted metabolite profiling/untargeted metabolomics** aims at the relative quantitative, non-selective, and universal analysis of all or as many metabolites (known or unknown) as possible in a given biological sample. The approach often does not have an *a priori* question, and therefore is hypothesis-generating. Accordingly, sample preparation is unspecific in order to capture as many metabolites as possible [1, 2, 4].

- **Metabolite fingerprinting** classifies samples based on their metabolite pattern ("fingerprint"). This screening approach does not differentiate between individual metabolites [1, 2, 4].

Metabolomics takes a snap-shot of the metabolism at a given time point, aiming to enable investigation of the physiologic state of an organism, diagnosis of diseases or metabolic disturbances, investigation of the influence of a certain treatment on an organism, or identification of marker candidates [2, 5, 6]. Especially in marker research, metabolite profiling, or untargeted metabolomics approaches are meant to identify potential metabolites of interest, which later might be applied for specific purposes using targeted analyses [7]. For these purposes metabolomics approaches are usually performed to compare metabolite profiles of groups (treated vs. control; healthy vs. diseased) and to apply statistical methods to identify differentiating metabolites [2, 5, 6]. Another important application of untargeted metabolomics approaches is to improve our understanding of the entire living organism at the molecular level (systems biology), meaning the biochemical and biological mechanisms in a complex system, together with the other "omics" approaches such as genomics, transcriptomics or proteomics [2]. Metabolomics is at the endpoint of the "Omics cascade", reflecting "what has happened and is happening", and thus is closest to the phenotype [2]. Therefore, it enables the most functional insight into the biological processes in a living organism, and the influence of or the response to stimuli are amplified in comparison to genomics, transcriptomics and proteomics [8].

According to the Human Metabolome Database (HMDB, version 4.0), the number of detected, expected and predicted human metabolites totals more than 114 000 (detected metabolites: ~21 700, expected metabolites: ~82 000; predicted metabolites: ~9 500) [9]. The metabolome consists of chemically highly diverse compounds with a wide range of polarity, solubility, and volatility – including but not limited to carbohydrates, secondary natural products, lipids, amino acids, terpenoids, steroids, aromatic compounds, and many more [10]. Additionally, these metabolites occur in a wide concentration range (7-9 orders of magnitude, pmol-mmol) [11]. Therefore, it is impossible to cover the whole metabolome with just one analytical technology [2, 3, 8]. With the aim to detect as many metabolites as possible, analytical technologies should fulfill the following criteria: non-selective, universal, robust,

sensitive, suitable for high-throughput analyses (especially for large studies), and sufficient separation performance [1-3].

1.1.2 Important aspects for metabolomics experiments

Experimental design

A typical workflow for a metabolomics study in nutritional research is depicted in Figure 1.1. The first and most important step is the experimental design, because the data can only be as informative as the experimental design allows [2, 12]. Therefore, unwanted variation (biological or technical/pre-analytical and analytical) has to be minimized and controlled. The choice of study design (observational vs. intervention study) and also its standardization (e.g. control and blinding, confounding factors, compliance etc.) are important [12]. For example, in the case of an intervention study, usually the influence of the high intrinsic biological variation due to many factors (among others sex, age, diet, life style, genetic background, hormonal changes, diurnal rhythms) has to be limited, while in the case of an observational question often the broad variation of a population is of interest. Generally, dependent on the study design, in order to obtain valid and statistically significant results for the investigated question, a large number of individual samples/participants is desirable [2, 6, 10, 13]. For example, for observational studies, a minimum of 100 participants is necessary to cover a broad population, while a minimum of 5-10 well characterized participants might be sufficient for an intervention study (better power calculation). Another relevant decision is the choice of the sample matrix for the investigation. Metabolomics analyses have been performed with many biological matrices and body fluids, such as urine, blood plasma or serum, feces, saliva, sweat, breath, or tissue samples. Attention should be paid to the following points: i) invasiveness of sample collection (e.g. urine vs. tissue) [14]; ii) biological question, availability of metabolites in a certain matrix (e.g. volatiles in breath vs. blood); iii) interference of matrix components with analytical technology (e.g. anticoagulants for plasma) [12]; iv) containers for samples and storage of samples until analysis (e.g. breath vs. urine) [1, 2, 14]; v) necessity of normalization (pre- or post-analytical; e.g. in the case of differentially diluted urine samples due to maintenance of water homeostasis) [12]; vi) influence of the background, for example macromolecules interfering with analysis (e.g. breath vs. blood) and vii) the sample preparation in general.

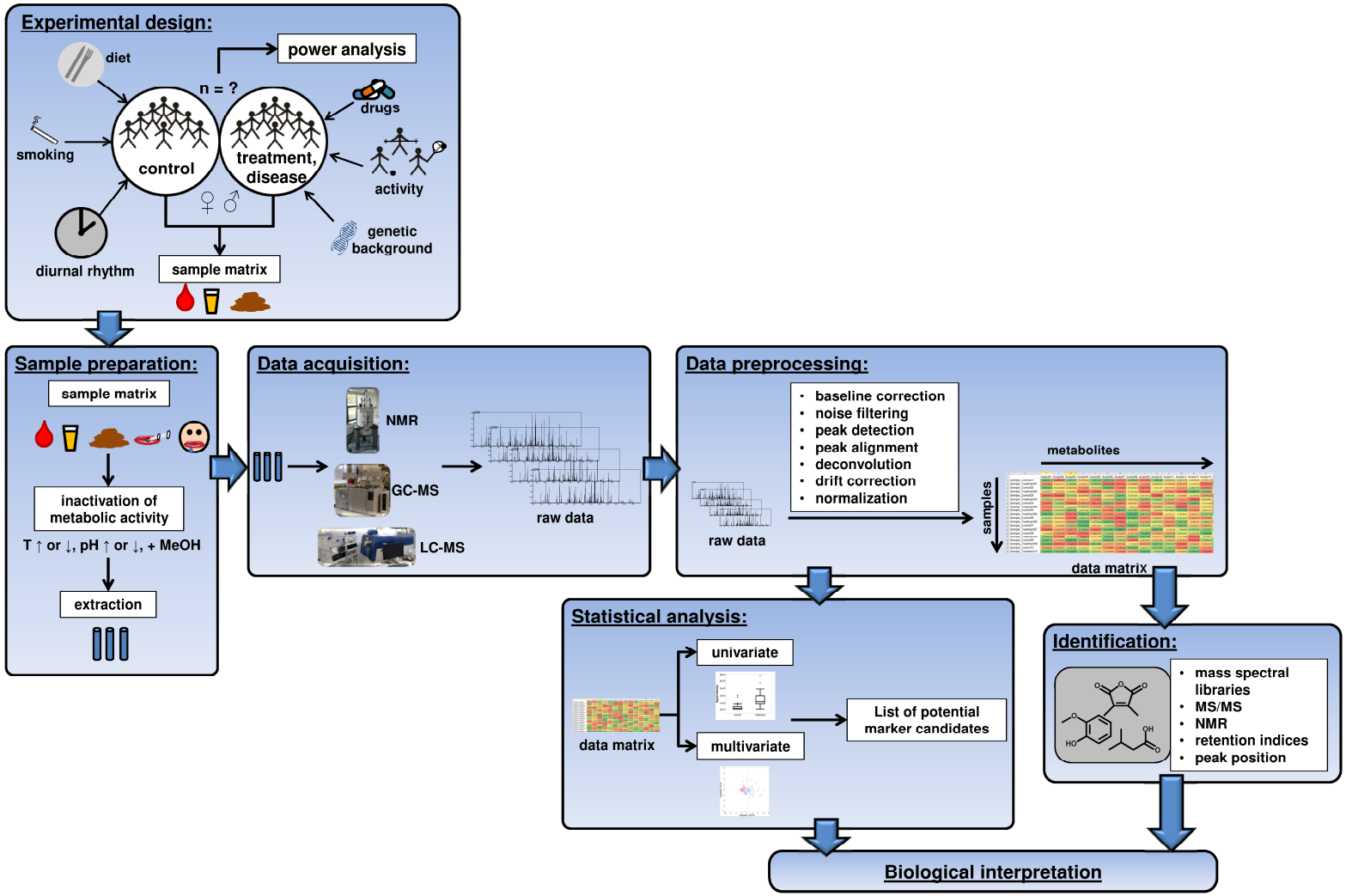


Figure 1.1. Typical workflow for human metabolomics experiment including experimental design, sample preparation, data acquisition, data preprocessing, statistical analysis, identification and biological interpretation (modified from [15]).

Sample preparation

Some of the stated points are also relevant for the next step in metabolomics experiments, the sample preparation. The sample preparation aims to lose as few metabolites as possible from a complex matrix, to avoid degradation or changes in metabolite levels, and to receive a sample that is compatible with the chosen analytical technology [2, 3]. Degradation of metabolites via enzymatic and non-enzymatic reactions have to be prevented using an inactivation or quenching step (e.g. freezing with liquid nitrogen, adding cold methanol, or adding acid and boiling) [2, 6, 8, 10]. Interferences by components of the matrix should be minimized; therefore, for example blood plasma or serum samples are usually subjected to a protein precipitation. Typically applied sample preparation methods are dilution, liquid-liquid extraction (LLE), solid-phase extraction (SPE), solid-phase micro extraction (SPME) or microwave-assisted extraction. The solvent applied for the extraction process, e.g. in the case of LLE, has a huge influence on the coverage of the metabolome and can introduce a bias with respect to some metabolite classes (polar vs. unpolar solvents) [2, 13]. To minimize pre-analytical variation due to sample collection, storage and preparation, all samples have to be treated consistently throughout the experiment [13].

Data acquisition

The three main methodologies applied in the untargeted metabolomics field are nuclear magnetic resonance spectroscopy (NMR), and mass spectrometry (MS) either coupled to liquid chromatography (LC) or gas chromatography (GC). Other more rarely applied technologies are for example direct injection MS or capillary electrophoresis coupled to MS (CE-MS) [2]. Each of these technologies has its own advantages and disadvantages with respect to its application in metabolomics. NMR enables i) a fast and non-destructive sample preparation with high-throughput; ii) a fast, highly reproducible, robust and non-discriminating quantification of metabolites; and iii) includes information on chemical structure. However, sensitivity and coverage of metabolites (only middle to highly concentrated metabolites detectable) is limited and identification based on chemical shifts of the signals is difficult in complex mixtures [2, 15]. In contrast, MS based methods allow a highly sensitive and selective detection and identification, especially in combination with a separation technology such as GC or LC. Disadvantages include a destructive and more complex sample

preparation and clean-up and also the possibility of discrimination against some metabolites or whole substance classes [2, 15]. While LC-MS enables the detection of the highest number of metabolites, GC-MS has the highest reproducibility with respect to retention time and mass spectra and is less susceptible to matrix effects and ion suppression than LC-MS [16]. However, GC-MS necessitates derivatization of metabolites to increase thermal stability and volatility [2, 16]. Each of these methods has an implicit bias with respect to metabolite coverage, therefore a combination of different methodologies enables a more comprehensive investigation of the metabolome [2]. To minimize intra-analytical variation, blank samples, internal standards (IS) and quality control (QC) samples are useful tools [2, 16]. Blank samples allow the assessment of artefacts due to derivatization reagents, solvents or in the case of chromatography due to the column. Internal standards and QC samples enable supervision of the data acquisition and performance of the equipment, and also the determination of the overall precision of the method. Additionally, QC samples can be used for correction of drift and batch effects [12]. Randomization of the sample order during the measurement is recommended to minimize the introduction of bias due to instrumental drifts [12].

Data preprocessing

After the data acquisition the raw data has to be converted to a form enabling further data analysis (data preprocessing). The commonly high number of samples and the huge number of metabolites detected per sample result in highly complex and large data sets requiring dedicated software programs for data preprocessing. Typical steps during data processing include baseline correction, noise filtering, peak detection, automatic annotation, peak alignment, deconvolution, drift correction, or normalization. Each step has to be well-conceived because every individual step can introduce undesired and uncontrolled post-analytical variation and can have fundamental influence on the results of the statistical analysis [10, 12].

Statistical analysis

Usually, a combination of univariate and multivariate statistics is applied to distinguish between groups. In the case of univariate statistic (e.g. t-test, analysis of variance (ANOVA) or correlation analysis) each feature is analyzed individually, while for multivariate statistic (e.g. principal component analysis (PCA), partial least squares (PLS) or clustering) all features are analyzed simultaneously [15]. As a first

step, statistical analysis aims to identify interesting features or patterns in the data due to the biological context. Such biological relevant features or patterns might, in a second step, enable the generation of a hypothesis for further investigations [10]. Important aspects with respect to statistical analysis are: i) each statistical approach has certain assumptions (e.g. normality, homogeneity of variances); ii) potential confounding variables have to be taken into account (e.g. sex or age); iii) proper validation strategies for results [12]; and iv) the problem of multiple testing due to the large number of features in metabolomics experiments. If a large number of features are analyzed at the same time, the likelihood of coincidental false positive associations is increased [15].

Metabolite identification and biological interpretation

One of the biggest bottlenecks in metabolomics studies is the identification of metabolites. As a result of advanced analytical technologies more and more metabolites can be detected, however many of these remain unknown and cannot be easily and directly identified [12]. Therefore, usually the statistical analysis is carried out and only relevant metabolites are subsequently identified, if possible. Different approaches for identification are conceivable: i) extensive databases with spectral information; ii) MS/MS or NMR experiments for structure elucidation; or iii) in the case of chromatographic separation the peak position together with retention indices (RI; only for GC) can give hints about the structure of a compound. For GC-MS in the case of electron ionization (EI), which produces highly reproducible mass spectra, extensive spectral libraries already exist. Spectral library databases such as the FiehnLib or the National Institute of Standards and Technology (NIST) libraries include mass spectra and alkane- or fatty acid methyl ester (FAME)-based RI's. Once relevant metabolites, distinguishing between a control group and a diseased or treated group, are identified, the last important step starts: the biological interpretation, which aims to find a mechanistic and contextual biological relationship between metabolites and the biological outcome.

The workflow of a human metabolomics experiment emphasizes that the whole metabolome cannot be covered by one sample preparation and/or analytical technology. Any sample preparation or analytical technology introduces a certain bias with respect to the coverage of the metabolome. Although sample preparation is usually as unspecific and general as possible, the high diversity in chemical

properties of the metabolites makes some losses unavoidable. For example, an extraction with a hydrophilic solvent often is only effective for polar substance classes and as a result excludes most unpolar substance classes. Another point is that some substance classes with very characteristic properties such as volatile organic compounds (VOC) need dedicated sampling approaches and analytical methods. With respect to the analytical technology, often a compromise between the detection of as many different metabolites and substance classes as possible and the resolution of some specific substances showing a high structural similarity, such as isomeric sugar compounds, has to be made. Therefore, for some selected metabolite classes or pathways, dedicated targeted comprehensive metabolite profiling approaches are desirable for highly resolved insights into the metabolome.

1.1.3 Technical aspects of GC-MS based metabolite profiling approaches

The development of capillary columns in GC enables the highest separation power and highest chromatographic resolution in comparison to other chromatographic techniques such as thin layer chromatography (TLC) or high pressure liquid chromatography (HPLC) [17]. Additionally, coupling with MS enables a high selectivity and sensitivity, and also the possibility to identify unknown metabolites. Moreover, the separation of chromatographically coeluting peaks using individual masses is achievable. Therefore, GC-MS represents an indispensable tool for untargeted metabolomics and metabolite profiling approaches [17].

Gas chromatography (GC)

The basic principle of GC is the separation of an analyte mixture through partition between a mobile gaseous phase (typically Helium) and a stationary phase (columns with different characteristic properties). Therefore, analytes have to be transferable to the gaseous phase (i.e. sufficient vapor pressure) and thermally stable. While volatile metabolites (such as many ketones, aldehydes, esters, furan and pyrrole derivatives, sulfides, isocyanates, isothiocyanates or hydrocarbons) can be directly analyzed, semi- and non-volatile metabolites (such as mono- and disaccharides, sugar phosphates, amino acids, fatty acids, small peptides, long chain alcohols, inositols, amines, amides, alkaloids, or organic acids) have to be derivatized to increase volatility and thermal stability [12].

VOC are important constituents of the metabolome [2, 18]. With respect to sample preparation and injection of volatile metabolites dedicated approaches are necessary. Conventional sample preparation approaches such as extraction or SPE are not feasible due to incomplete extraction or losses during concentration steps, and additionally, solvents might interfere with analyte elution (i.e. solvent front) [2]. Therefore, headspace (HS) techniques are applied, such as static or dynamic HS or HS-SPME [17]. Headspace refers to the gaseous phase above solid or liquid matrices [19]. In the case of static HS-GC, an aliquot of the equilibrated headspace above a sample matrix is injected into the GC, while in dynamic HS-GC or "purge-and-trap", the sample is continuously and exhaustively extracted before desorption in the GC [17, 19]. Another simple and cost-effective HS sampling techniques, which is often applied in metabolomics is SPME, which integrates sample collection, extraction and concentration in one solvent-free step [18, 20]. A fused-silica fibre situated in a hollow needle is coated with a stationary phase, resembling a syringe assembly. The fibre can either directly be immersed in a liquid sample (direct immersion) or in the HS above the sample (HS-SPME), during this time VOC are either adsorbed or absorbed on the stationary phase. HS-SPME is a complex multiphase equilibrium. In addition to the equilibrium between sample and HS (similar to static HS-GC), there is the equilibrium between HS and stationary phase of the fibre [17, 18]. Similar to static HS the following parameters are important for method optimization: extraction temperature and time, salting-out effect, agitation of sample, equilibration time, phase ratios (sample/HS; HS/stationary phase), film thickness of stationary phase (influences equilibration time and sensitivity) and interactions between analytes and matrix [17, 18]. To achieve high precision and reproducibility, a strict standardization of sample preparation and conditions is necessary, especially if extraction is performed under non-equilibrium conditions [18]. Untargeted metabolomics and metabolite profiling approaches aim to detect as many compounds as possible, therefore stationary phases with a low selectivity are desirable [21]. Many GC-based metabolomics studies recommend a mixture of DVB/CAR/PDMS (divinylbenzene/ carboxen/polydimethylsiloxane) as stationary phase for the fibre to detect a variety of VOC of a broad range of volatility and polarity [21-23].

Most primary metabolites such as amino acids, organic acids, sugars, lipids and alcohols require derivatization to increase volatility and thermal stability before GC-

MS analysis. Derivatization should be fast, simple, efficient, reproducible and if possible automated via an autosampler to reduce batch effects (e.g. degradation or variability in ratios of partially derivatized metabolites) [12, 17]. A number of derivatization methods exist such as silylation, acylation, alkylation/esterification and many more. However, most GC-MS based untargeted metabolomics or metabolite profiling methods apply a two-step derivatization including an oximation followed by silylation [2, 12, 16, 17]. Oximation, either performed with hydroxylamines or alkoxyamines, prevents cyclization of sugar isomers, decarboxylation of α-ketocarboxylic acids, and keto-enol tautomerism. Reducing sugar compounds equilibrate in aqueous solutions between their open-chain and cyclic α- and β-semiacetal form. If only silylation is performed, this results in six and more derivatives (Figure 1.2) [2, 12], which in turn leads to a more complex chromatogram and separation problems. Therefore, oximation or methoximation is performed, arresting reducing sugars in their open-chain form and leading to only two isomers (*syn* and *anti;* Figure 1.2).

Following methoximation, typically a trimethylsilylation is performed, often using reagents such as *N*-Methyl-*N*-(trimethylsilyl)trifluoroacetamide (MSTFA) or *N,O*-Bis(trimethylsilyl)trifluoroacetamide (BSTFA). During silylation active hydrogen in functional groups (such as OH, COOH, NH or SH) is replaced by trimethylsilyl(TMS)-groups. However, silylation is a reversible process, even traces of water or humidity might cause degradation. Consequently, taking care of evaporation of extracts before derivatization and close systems while handling derivatization reagents (i.e. working under inert gas) is necessary [2, 12]. Limitations of derivatization procedures are: i) it causes longer and more complex sample preparation and can introduce additional variance; ii) in the case of some compounds degradation or rearrangement reactions are possible (e.g. arginine to ornithine or glutamate to pyroglutamate); iii) especially for amino acids a partial silylation can occur, leading to multiple derivatives; iv) due to reactions of the silylation reagent with itself or solvents, artifacts can be formed; and v) it can lead to altered compound compositions in the samples (e.g. via contamination or faster/selective reaction with certain substance classes) [2, 12, 17, 24]. However, derivatized compounds show more distinct MS spectra in comparison to the corresponding underivatized compounds and an improved sensitivity and selectivity is observable [2, 17]. An important reason, why trimethylsilylation with and

without oximation/methoximation are mainly used is that extensive EI-MS spectral libraries exist (e.g. NIST or FiehnLib).

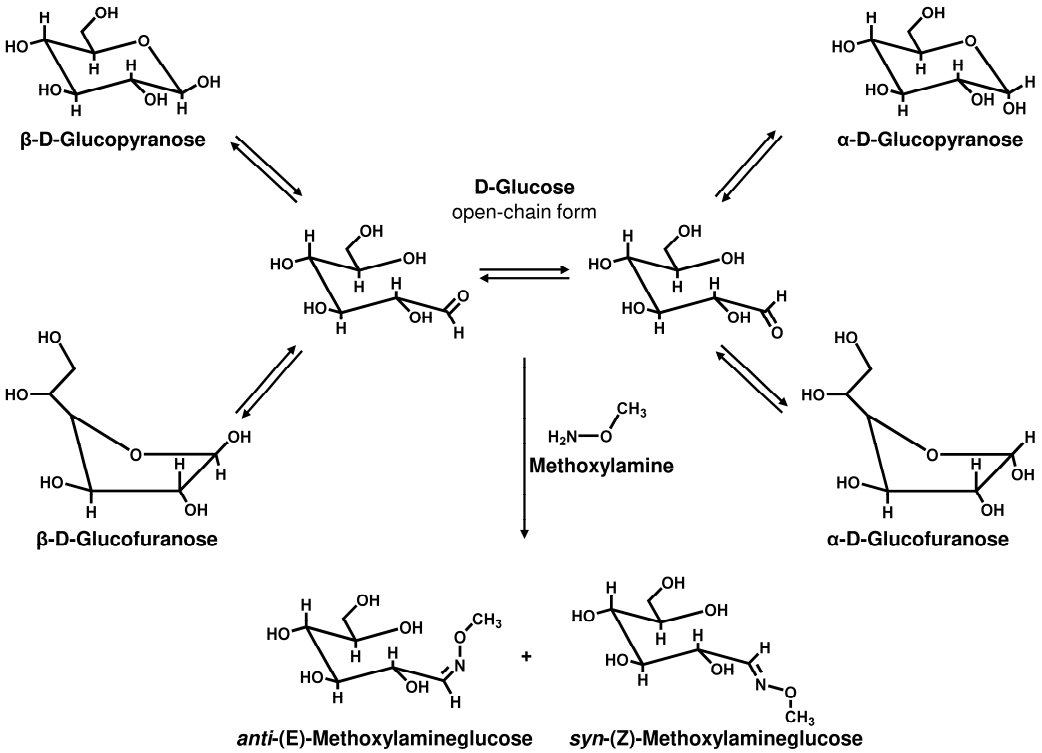


Figure 1.2. Mutarotation of glucose in aqueous solution and methoximation (modified from [12]).

Coupling with mass spectrometry

The combination of GC with MS is very advantageous and additionally, the coupling of capillary columns with MS is easily realized [17]. For MS, ions are generated from molecules in the ion source at temperatures between 200-300 °C, then separated according to their mass-to-charge ratio (*m/z*) in the mass analyzer and finally detected in the detector; all this happens under high vacuum conditions (10^{-7}-10^{-5} mbar) [12]. Two regularly applied MS techniques in metabolomics are quadrupole MS (qMS; Figure 1.3) and time-of-flight MS (TOF-MS) [2]. The qMS uses four parallel metal rods with direct and alternating currents as mass filter, so that at a given setting and time period only ions with a certain *m/z* have a stable trajectory and can pass the quadrupole rods towards the detector. The current is continuously changed over time to analyze the desired mass range; this results in slow scan rates dependent on scan speed and selected mass range [2, 12, 17]. For TOF-MS, ions are

accelerated in an electric field; their velocity and the transit time through a drift region depend on their mass. Therefore, high acquisition rates (up to 500 spectra/s) are possible and the whole mass range is acquired simultaneously; it can be operated both in nominal and high mass resolution [2, 12]. However, TOF-MS are comparably sensitive and cost-intensive [17]. In the last years, Q-TOF-MS and Orbitrap are emerging MS to be coupled with GC. Q-TOF-MS permits to determine parent and fragment ions along with their isotopic abundance with high mass accuracy and thus, is especially useful for identification purposes [17]. Similarly, Orbitrap enables even higher mass resolution and accuracy, also providing information on MS/MS spectra, however acquiring data at high resolution will lead to reduced number of scans/peak [12].

qMS, which were utilized in this thesis, can be operated both in scan mode and selected ion monitoring (SIM)-mode. Untargeted metabolomics approaches mainly operate in scan mode due to the possibility to record more comprehensive information for metabolites and therefore better identification. SIM mode is an interesting alternative for targeted or metabolite profiling approaches aiming to identify and quantify targeted substances or substances of certain substance classes [12]. Advantages of SIM mode are i) lower limits of detection due to longer dwell times of selected *m/z*; ii) better separation of metabolites and especially better separation from matrix compounds; and iii) better accuracy and precision due to more data points, which are equally distributed over the peaks [12, 14]. Nonetheless, SIM mode necessitates prior knowledge of the measured compounds (i.e. characteristic *m/z* for certain metabolites or whole substance classes) and is more of a targeted nature or semitargeted, if the monitoring of a whole substance class is the aim [14]. In consequence, the coverage of the metabolome is naturally restricted.

In general, ions for GC-MS measurement are mostly generated by EI or chemical ionization (CI) [2, 12, 17]. EI, a hard ionization method with extensive and highly reproducible fragmentation patterns, is the ionization technique of choice [2, 4, 12, 17]. Gaseous compounds are bombarded with electrons and as a result radical cations are formed, which are then extensively fragmented. 70 eV are used for EI, even though for the ionization of most molecules 10-15 eV would be enough, however the ionization efficiency and in consequence, the sensitivity would be limited [12, 17]. The

resulting highly reproducible and extensive fragmentation patterns can be used for identification of metabolites using existing commercial EI-MS spectral libraries, some of which already include mass spectra and RI's of more than 99 000 compounds (NIST17, National Institute of Standards and Technology, Gaithersburg, MD, US). However, the following limitations exist:

- missing molecular ions complicate identification of unknown compounds not listed in spectral databases
- fragmentation ions are often unspecific and difficult to interpret
- structural isomers very often exhibit highly similar EI spectra necessitating chromatographic separation for an unambiguous identification (e.g. in the case of sugar compounds) [12, 17]

The additional use of CI, a soft ionization technique, might enable overcoming some of the limitations with respect to identification. In the case of CI, the energy transmitted to the analyte molecules is low, therefore the formed molecular ion is relatively intensive, and less fragmentation occurs [12]. However, CI results in less efficient ionization and a loss in sensitivity. Additionally, spectra obtained in different labs are not as comparable and therefore, building databases of mass spectra is complicated. Nonetheless, the more intensive formation of the molecular ion is advantageous for the calculation of the molecular formula of unknown compounds and, together with fragmentation patterns obtained by EI-MS, might enable a more reliable structure elucidation. However, in the case of structurally highly similar isomers (e.g. sugar compounds), even this will not enable unambiguous identification. Such challenges might be overcome by using enhanced separation capacity, as it can be achieved for example with comprehensive two-dimensional gas chromatography (GC×GC).

Comprehensive two-dimensional gas chromatography (GC×GC)

The fundamental principal of GC×GC is the separation of all analytes on two columns with different stationary phases. The peak capacity, a tool to evaluate chromatographic separation and column efficiency, is increased in a multiplicative way in comparison to one-dimensional GC ($n_{total} = n_1 \times n_2$) [17]. A comprehensive two-dimensional separation has to fulfill the following criteria according to Giddings *et al.* [25]:

- All analytes of a sample have to undergo two or more essentially different/independent separation mechanisms (orthogonality criterion).
- Analytes that were separated in the first dimension (^{1}D), have to remain separated in the second dimension (^{2}D) (modulation criterion).

The set-up of the GC×GC-qMS systems applied in this work are depicted in Figure 1.3.

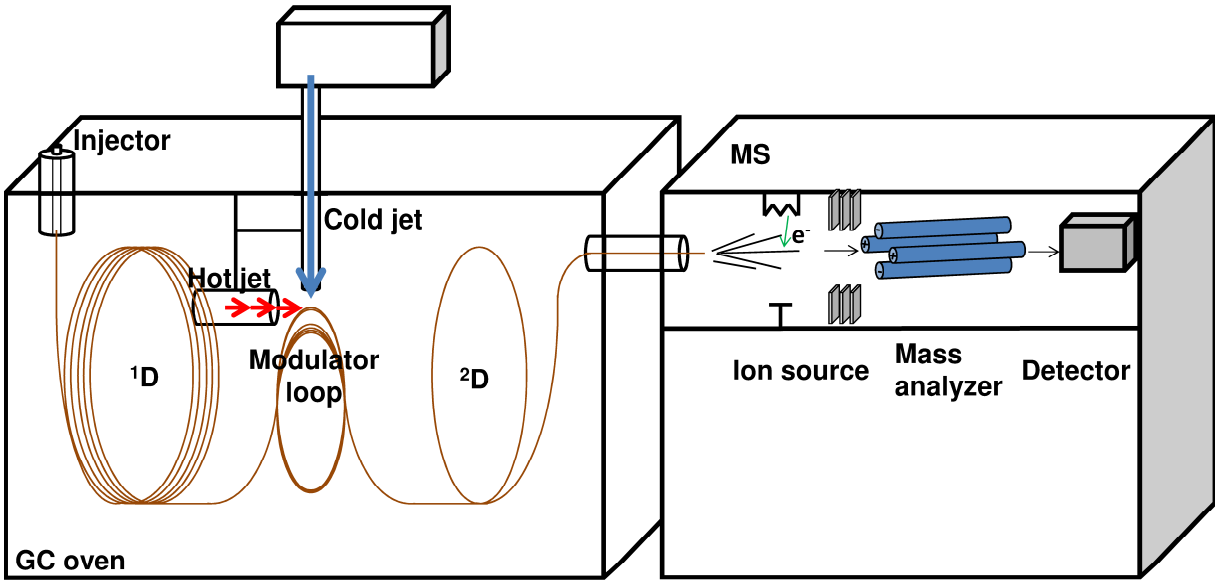


Figure 1.3. Set-up of the GC×GC-qMS systems applied in this work.

In general, two columns with different stationary phases are serially connected, with the so-called modulator lying in between. To fulfill the orthogonality criterion, the choice of stationary phases in the ^{1}D and ^{2}D is essential. Often a column with an unpolar stationary phase is chosen for the ^{1}D-separation, which is based on vapor pressure, and a column with a polar or medium-polar stationary phase is chosen for the ^{2}D-separation, where additionally polar interactions with the stationary phase, such as hydrogen bridge bonds or dipole-dipole interactions, play a role [12, 17]. Non-orthogonal column combinations with a polar stationary phase in the ^{1}D and an unpolar one in the ^{2}D are less common, but can be advantageous for certain applications, such as for polar and/or ionogenic analytes or for VOC analysis [12, 26-28].

The modulator fractionates and focusses the eluate from the ^{1}D-column, transfers and reinjects it for further separation on the ^{2}D-column [12, 28-31]. In general, thermal and differential flow modulators can be differentiated:

- **Thermal modulators** (jet-based cryogenic modulator, utilized in this work): Part of the column is placed in the air streams of two jets (cold and hot jet). The cold jet continually emits a cold air stream and focusses the analytes in a narrow band. Possible modulator gases are nitrogen, carbon dioxide or compressed (dried) air; the choice depends on volatility of analytes and cost. The hot jet emits fast, hot pulses to mobilize and reinject the analytes on the separation length of the ^{2}D-column [12, 17, 28, 32].
- **Differential flow modulators**: Utilizes a six-way-valve and a second flow controller and so far is rarely applied in praxis, because the technical development and optimization of differential flow modulators is still emerging. The eluate of the ^{1}D is collected in a loop or chamber and after switching the valves transferred and reinjected fast on the ^{2}D-column via the additional flow controller unit [12, 17, 28, 32].

Thermal modulators are preferred, because in addition, they enable effective focusing of analytes in narrow bands, leading to narrow peaks with optimal peak shape and increased sensitivity [12, 17, 28, 29, 32]. However, the technical realization and hardware is more complex and cost-intensive in comparison to flow modulators. Especially, in the case of highly volatile analytes liquid nitrogen is necessary as modulator gas. The duration of each fractionation of the ^{1}D-eluate, the so-called modulation period, is typically between 2-8 s. The separation on the ^{2}D-column has to be completed within the modulation period or, in other words, before the next fraction is reinjected. *Wrap-around* occurs if analytes elute from the ^{2}D-column one modulation period later than they were reinjected on the ^{2}D-column. The reverse effect, so-called *breakthrough*, describes the effect of highly volatile analytes, which are not sufficiently fractionated in the modulator.

Peaks eluting from the ^{1}D-column have peak widths between 6-25 s and have to be modulated at least 3-4 times to fulfill the modulation criterion and ensure that the separation of the ^{1}D is maintained. As mentioned above, the modulation creates narrow peaks in the ^{2}D with peak base widths of 50-200 ms enabling improved signal-to-noise-ratios and higher sensitivity [12, 17, 28, 29, 32]. To accurately reconstruct these narrow peaks, a minimum of 8-15 data points is required, which necessitates fast detectors [12, 17, 33]. Detectors fulfilling this requirement are the flame ionization detector (FID), the electron capture detector (ECD) or the MS detector, which

enables additional structure elucidation. Especially TOF-MS, with data acquisition rates of up to 500 Hz, are often applied. However, qMS systems with fast scanning properties are cost-efficient and robust alternatives, although the mass range is limited [31, 33].

Processing of GC×GC-MS data is more complex in comparison to one-dimensional GC-MS data and dedicated software and algorithms are needed. Each chromatogram consists of numerous, short, consecutive ^{2}D-chromatograms, which have to be rearranged for the usual 2D-depiction as a contour plot (Figure 1.4). Additionally, the modulation process leads to multiple peaks per analyte [12], which have to be recombined. A huge data volume is the result of measuring a few hundred analytes using a GC×GC-MS approach with each analyte having multiple modulation peaks.

Two different approaches to handle GC×GC-MS data have been described so far:

- **Peak-feature based approach**: For this approach the peak information (e.g. retention time, RI, peak height/area, mass spectral information, library matching, etc.) is extracted similarly to the traditional one-dimensional data processing. The demodulation, meaning the combination of multiple modulation peaks belonging to the same analyte, is performed with dedicated software. The final result is one textual peak list per raw data file with considerably smaller size, which can be used for further processing steps such as drift/batch correction and alignment [12].
- **Peak-region feature approach**: For this approach the two-dimensional (2D) contour plot is utilized to extract information based on methods developed for pattern recognition, which have been implemented in commercial software packages (GC Image, GC Image LCC, Lincoln, NE, US). The information of 2D-peaks (i.e. 2D-peak contour) and 2D-chromatographic regions, which describe a chromatographic region not necessarily including only one compound, can both be utilized to generate a template describing the pattern of all detectable compounds. This template includes mass spectral information, information about the 2D-retention search space and the absolute and relative position within the pattern of detectable 2D-peaks and 2D-chromatographic regions. It can be utilized both to integrate raw data files and

for alignment purposes. Additional steps such as drift/batch correction may be performed subsequently [12].

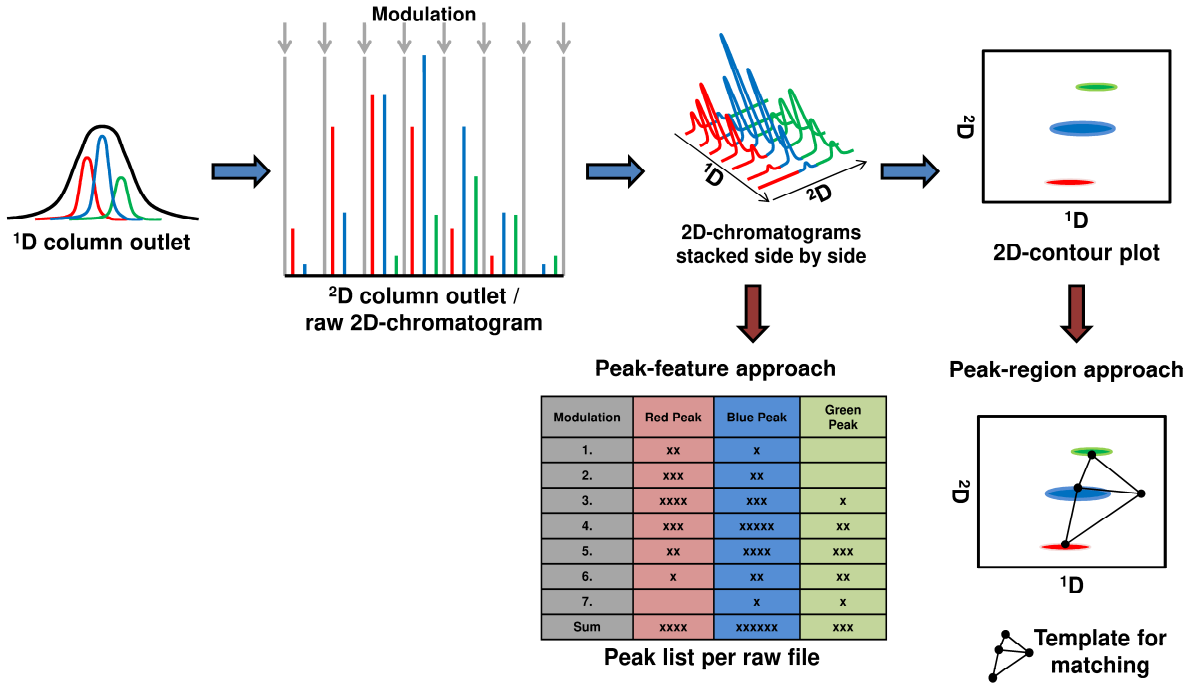


Figure 1.4. Raw data processing of GC×GC-qMS files, including both peak-feature and peak-region based approaches (modified [34]).

1.2 Applications of metabolite profiling approaches in nutrition and health

1.2.1 Investigating associations between nutrition and health – general aspects

There is an increasing incidence of diet-related diseases worldwide, which will become even more relevant in the future [35]. On the one hand, an unbalanced diet can promote or even cause a disease or metabolic disorder, and on the other hand, the general health status or even existing diseases can be positively influenced through nutrition and the intake of specific foods, food constituents or essential nutrients [36-39]. An important goal in nutritional research is to estimate relative risks between dietary factors and the occurrence of diseases [38]. Further, especially the early diagnosis of diseases or metabolic disorders is crucial because their onset can effectively be prevented or at least delayed by preventive measures during early developmental stages. For example, in the case of type 2 diabetes mellitus, diet and life style changes during the prediabetic state are successful in preventing the onset of diabetes.

Observational studies are employed to examine relationships between nutrition and diseases or health status [38, 39]. For these investigations, the measurement of the dietary intake is essential. Dietary intake data is then correlated with clinical parameters and/or the occurrence of diseases such as cancer, cardiovascular disease or diabetes to find potential associations. However, a correlation does not imply a causal relation. Consequently, intervention studies are often performed in addition to substantiate the influence of dietary factors (e.g. polyphenol-rich, Western diet, Mediterranean diet, essential nutrients or special food components) on the human body and health status [40]. However, three main limitations lead to substantial variation, which in turn complicates finding unambiguous and reliable associations between health status and nutrition:

- **Uncontrollable conditions and incomplete standardization:**
  Uncontrollable conditions and also incomplete standardization in human studies complicate the interpretation of research findings in nutritional studies [41]. Uncontrollable conditions include factors, such as genetics (polymorphisms, phenotypic variations), environment, life style, age, and sex. All of these factors can have a considerable influence on the individual

response to a certain food and therefore, might in part confound results [41, 42]. Yet, at least in part, some of these confounders, e.g. sex and age, might be accounted for in the applied statistical approach. Incomplete standardization with respect to study design variables (e.g. compliance, additional food intake, wash-out and run-in phases during studies) or, depending on the research question and study design, the incomplete characterization of the study population (e.g. genetic polymorphisms or individual gastrointestinal microbiota) can have a considerable influence on the outcome of a study [36, 38, 41, 42]. It is not possible to account for incomplete standardization, e.g. for compliance, in statistical approaches. The resulting high variation necessitates large participant collectives to attain the statistical power to discover small effects and detect reliable associations between nutrition and health [38, 43].

- **Assessment of dietary intake:**

  Next to standardization and variations in phenotypes, dietary assessment methods add significantly to the variation of the data with considerable random and systematic errors [38, 39, 43]. In general, in dietary assessment, the actual diet, i.e. consumption within the last 24 hours, and the habitual diet, i.e. consumption over a longer time period (e.g. several months or a year), are differentiated. During epidemiologic studies, food frequency questionnaires (FFQ) are often applied to assess the habitual diet via the frequency of consumption and typical portion sizes, which is an easy and cost-effective way to assess diet [43]. 24 h recalls are structured interviews assessing the consumption of food and beverages during the last 24 hours, which uses picture booklets of typical portion sizes to estimate the consumed amount, like FFQ. Therefore, both are prone to errors in estimating the consumed amount, while in the case of weighed records, participants record the detailed amounts of consumed food and beverages in a diary. However, weighed records might influence and change the behavior of participants and compliance might be a problem due to the time effort. Retrospective methods, such as FFQ and 24 h recall, require a good memory, with respect to both choice and amount of food. While in the case of FFQ daily variations in consumption and possibly changed dietary habits during the study period may lead to erroneous data, particularly spontaneous snacks might be forgotten during 24 h recalls. Next to this unintentional misreporting, all self-reported dietary assessment methods

are prone to intentional misreporting, with respect to both choice and amount of food, due to social desirability and approval, e.g. underreporting in the case of perceived unhealthy food or overreporting for perceived healthy food [43]. The studied population might also have an influence on misreporting, e.g. it was observed that overweight participants are more prone to underreporting energy intake than others [39, 43]. Accurately assessing dietary intake is essential in investigating associations between diet and health status [44].

- **Complexity of the consumed food and its potential effects:**
  Other problems related to the traditional dietary assessment instruments are the details in the asked questions during a FFQ or 24 h recall. Equally important is the dependency on comprehensive food composition and nutrient databases to calculate the total energy intake, the nutrient intake, or the intake of other non-nutrients from the estimates of dietary intake [38, 39]. These food composition and nutrient databases are limited with respect to the enormous complexity of compounds occurring in food. While calculating the intake for macro- and even some micronutrients is possible, calculating other potentially bioactive compounds (such as secondary plant metabolites) is almost impossible due to incomplete data for such compounds in these databases [38, 45]. Additionally, neither FFQ nor 24 h recalls differentiate between different varieties (e.g. orange and blood orange) due to missing details in the asked questions. However, factors such as variety, environment (e.g. place of growth, seasonal changes, weather differences), processing (e.g. fertilization, cooking, baking, etc.) or storing can have considerable influence on food composition [45, 46]. Therefore, another source of error is due to the calculation based on average amounts of compounds in a specific food. Furthermore, the complexity of more than thousand compounds in a food item complicates both, the determination of associations of only one specific compound (e.g. a flavonoid) or of a certain diet (e.g. Mediterranean diet) with an observed health effect. Any diet consists of a whole variety of potentially bioactive compounds, all of which might be individually or together responsible for a health effect in a large number of biochemical pathways [42, 44, 47, 48].

Metabolomics and metabolite profiling approaches might facilitate the detection of associations between health status, diseases and nutrition by addressing and

circumventing some of the abovementioned limitations. In this respect, metabolomics and metabolite profiling approaches might support research in nutrition and health:

- by accurately and comprehensively describing the human metabolism and metabolic phenotypes, including the metabolic effect of individual genetic polymorphisms [36, 41, 49];
- by complementing traditional dietary assessment methods and using dietary intake markers [38, 40, 46];
- by uncovering associations between human nutrition, metabolism and health status [49] and consequently, by understanding mechanistic relations between nutrition, metabolism and health status [36, 37, 49, 50];
- and by comprehensively describing the components of the diet itself [40].

In general, metabolomics can help improving the so far incomplete knowledge and understanding of the metabolism itself, the alterations in metabolism in the case of diseases, the biochemical function of different metabolites and also the interaction between nutrition and health status [36, 42, 47, 49]. The discovery of new diagnostic biomarkers especially for pre-disease states might be facilitated by metabolomics. However, the implementation and success of the abovementioned points is not trivial due to the complexity and interplay of many factors and will take great effort in the coming years [36, 42, 47, 49].

The terms biomarker and marker are sometimes used synonymously and sometimes they are differentiated. Additionally, there are many different classification systems for biomarkers causing ambiguities concerning the use of these terms, their validation and subsequent application [50-53]. A summary of different definitions and classifications can be found in Supplemental Table S1.1. In general, a marker is reflecting or “marking” an exposure, a status, a function, an effect or a risk factor and therefore, is always applicable [53]. A biomarker, as defined by the WHO is a chemical, its metabolite, or the product of an interaction between a chemical and some target molecule or cell that is measured in the human body [54]. The WHO differentiates three subclasses of biomarkers [55]:

- **Biomarker of exposure:** an exogenous substance, its metabolite, or the product of an interaction between a xenobiotic agent and some target

molecule or cell that is measured in a compartment within an organism and can be related to exposure [54, 55].

- **Biomarker of effect:** a measurable biochemical, physiological, behavioral, or other alteration in an organism that, depending on the magnitude, can be recognized as associated with an established or possible health impairment or disease [55].
- **Biomarker of susceptibility:** an indicator of an inherent or acquired ability of an organism to respond to the challenge of exposure to a specific chemical substance [55].

Often the term biomarker (of effect) is used more specifically for surrogate endpoints in clinical settings to predict future events, and to indicate normal biological processes, pathogenic processes, or pharmacologic responses [44, 53]. Dietary or nutritional markers reflect an exposure to the intake of a food constituent, the entire food or maybe even a dietary pattern, and they are measured objectively in biological specimen [44, 48, 53]. Although the term biomarker is often used for dietary markers in the literature, a differentiation between biomarkers (clinical setting – biomarkers of effect) and markers of food intake and dietary exposure (biomarkers of exposure) is made in this thesis.

1.2.2 Markers for dietary intake

An accurate assessment of dietary intake is crucial in investigating associations between diet and disease, as both qualitative and quantitative changes in food can strongly influence the risk for diet-related diseases [50, 51]. Due to concerns with respect to validity and quantitative accuracy of the traditional dietary assessment methods, dietary markers detected in body fluids like urine, blood or even hair are of utmost interest. In future, dietary markers might facilitate a more objective and precise measurement of dietary intake, especially if used in conjunction with traditional dietary assessment [56]. In this respect, dietary markers are not meant to replace, but to complement the traditional dietary assessment methods. Most foods consist of thousands of compounds, which theoretically all represent potential markers of intake of the respective food. Some of the compounds will be excreted unchanged from the human body dependent on the consumed amount [39, 45], while others are transformed via phase I or II metabolism (e.g. oxidation, glycosylation, glucuronidation or sulfonation) resulting in an even higher number of metabolites [40, 41].

Next to the objective measurement of the true intake itself [44, 48], dietary markers might be used to ensure and monitor compliance of participants in a long-term intervention study or in patients during treatment (e.g. abiding dietary instructions) [44, 48], as a tool to assess biological effect of nutrition and susceptibility to nutritional effects [50] or to assess and calibrate the measurement error and, therefore validating results from traditional dietary assessment instruments [38, 40, 44, 46]. Application of validated and established markers, such as doubly labelled water for total energy intake or 24 h urine nitrogen for protein intake, revealed considerable differences to total energy and protein intake measured with traditional dietary assessment methods. In these studies, comparing traditional dietary assessment and established markers, the uncertainties connected with traditional dietary assessment methods lead to weaker observed effects and a loss of statistical power, explaining that often no relation between diet and health status is detected or that results are contradictory [38, 43]. So far however, only a few dietary markers are established and validated. A few examples of marker candidates for dietary consumption are listed in Table 1.1. Therefore, new dietary markers have to be identified and validated in the future to cover a broader spectrum of food intake, provide a deeper understanding of metabolic dynamics and enable the detection of associations between nutrition and health status [38, 45-47, 50].

In general, it is important to perform both, marker discovery and marker validation. Two possible strategies for marker discovery exist: hypothesis-driven or data-driven [44, 48, 72]. Most of the markers that are already in use were discovered using the hypothesis-driven approach, which necessitates *a priori* knowledge of the potential marker. Markers are usually either compounds of the food itself, their metabolites after digestion, absorption and/or metabolism or endogenous metabolites that increase as a result of altered metabolic pathways (e.g., serotonine metabolism after alcohol intake [73]) [48]. Food components can be metabolized in three different ways [48]:

- digestion of larger food components into simple nutrients in mouth, stomach and small intestine, so that they can be absorbed (typically lipids, polysaccharides, proteins);

- further transformation by host tissues (liver and kidneys) through biotransformation in phase I and II metabolism (xenobiotics such as polyphenols, alkaloids, carotenoids, volatiles, flavors and so on);
- processing by gut microbiota in the large intestine.

Table 1.1. Examples of dietary marker candidates.

Food or food group	Marker compound/s	Matrix	References
protein intake	24 h nitrogen	urine	[38]
total sugar intake	fructose+sucrose	urine	[57-60]
alcohol	ethylglucuronide, ethylsulfate, ethanol	urine, blood, breath	[61-63]
wholegrain wheat and rye intake	alkylresorcinols	urine, blood	[64]
citrus fruit, particularly orange	proline betaine	urine	[65]
red meat	*O*-acetylcarnitines	urine	[66]
chocolate (solid or drink)	6-amino-5-(*N*-methylformylamino)-1-methyluracil, theobromine, 7-methyl-uric acid	urine	[67]
coffee	trigonelline, cyclo(isoleucylprolyl)	urine, blood	[68]
Allium vegetables (onion, garlic)	*N*-Acetyl-*S*-(2-carboxylpropyl)cysteine	urine	[69]
dairy products	galactose, lactose	urine	[70, 71]

Increasing databases (such as FooDB, HMDB or Phenol-Explorer) will be important ways to find information on food composition and typical metabolization in future. And subsequently, allow building hypotheses for discovery of potential marker compounds.

On the other side, the growing numbers of metabolomics investigations, which show changes in metabolite profiles as a result of the diet, enable data-driven approaches for marker discovery [44, 48, 72]. For this, either samples from controlled dietary intervention studies or samples from cross-sectional or longitudinal studies can be used [48, 72]. In the case of controlled dietary interventions, either a single meal/food is given (acute study) and samples collected postprandially for up to 48 hours or a meal/food is consumed for a longer period (from a few days up to six months) and samples from the beginning and end of the period are compared e.g. with a control. In the case of a cross-sectional or longitudinal study, participants are grouped into

low and high intake according to self-reported intake and metabolites are compared via statistical approaches or correlation analysis with self-reported intake and statistical approaches are performed. On the one hand, observational studies have the advantage of more robust markers due to the availability of a background diet, however, on the other hand they have the disadvantage of using self-reported intake for the grouping of participants, which is prone to misreporting [48, 72]. One of the major disadvantages of the data-driven approach via untargeted metabolomics is that a marker can only be used, if it is possible to elucidate its structure. This is one of the major limitations of metabolomics approaches in general (section 1.1.2).

More and more markers are discovered nowadays by metabolomics approaches (see tables in [48]). However, marker validation, which is necessary before application of any marker in an epidemiologic study or in a clinical setting, has been performed rarely, if at all [44, 52, 72, 74]. This is due to the fact that so far no standard procedure or guideline for validation of dietary markers exists. In the FoodBAll (Food Biomarker Alliance) Consortium, a multidisciplinary team developed a first scheme for validation of dietary markers [74]. Validation in an independent study is a crucial step and is meant to ensure that a marker is enabling appropriate results for its intended purpose [52, 72, 74]. Dragsted *et al*. [74] proposed eight criteria for validity of markers of food intake, the first five criteria relate to biological and interpretative validity and the last three criteria relate to the validity of the applied analytical method. Variability in the consumed amount of food, determined via a dietary marker, can be a result of differences in absorption, distribution, metabolism and excretion due to factors such as age, sex, genetics, life style (e.g. smoking, alcohol consumption, drugs), health status, background diet, individual gut microbiota or variable amounts of the marker compound or its precursor in the food itself [44, 50, 74]. Additionally, the stability of the marker compound might be influenced by sample collection, processing or storage [44, 50, 74]. The aim of the eight criteria for validation is to ensure a more objective and precise measurement of the markers of intake. In the following section the eight criteria are described and explained as suggested by Dragsted *et al.* [74]; a deeper discussion of their applicability, advantages and disadvantages can be found in the general discussion section:

- **Plausibility (biological and chemical)**:

  Plausibility means that a causal explanation for the increase of a putative dietary marker after intake of a certain food or food group exists [44, 53, 74]. Next to the occurrence of a putative marker in the targeted food, knowledge on variability (e.g. based on variety, growth conditions, etc.), on food chemistry (e.g. process-derived components) as well as bioavailability, digestion, distribution, metabolism and excretion play an important role for the evaluation of plausibility [44, 46, 48, 74]. In this respect, food metabolite profile databases are highly desirable, but at the moment only in a limited way available. Databases such as FooDB are being set up at the moment and will be important resources in the future [9]. Dragsted *et al.* [74] include in this criterion also the question of specificity of a marker, meaning for example the occurrence and amounts of the marker metabolite in other food and also the amounts and frequency of consumption of such food.

- **Dose-response**:

  For the quantitative estimation of the intake, the excretion of the putative marker should be a fixed proportion of the intake [44, 74]. The dose-response relationship is important to determine the potential range of intake, from limits of detection due to high baseline levels or low bioavailability to saturation effects (e.g. vitamin C) [44, 48, 72, 74]. Evaluation of dose-response has to be performed in a controlled intervention study with a well-defined food and different amounts of intake, where no background diet could possibly influence the excreted amount of the putative marker [44, 48]. Nonetheless, a cross-sectional study or longitudinal study with different intakes can be used as well, but with less certainty (due to bias through background diet or usage of self-reported intake).

- **Time-response**:

  Time-response addresses the kinetics (absorption, distribution, metabolism and excretion) of the putative marker in the human body with respect to its biological half-life. And consequently, whether the putative marker enables the detection of intake over hours, days, weeks, months or even years (short- vs. medium- vs. long-term markers) [74]. Both, the time-response after one meal, and also the time-response after repeated meals (i.e. reproducibility of response, e.g. accumulation effects, cumulative increase of excretion due to

induced or inhibited enzymes or altered excretion) are important to be assessed within controlled intervention studies to decide on the best sampling time and biological sample [74]. Most often highly polar compounds have a short physiological half-time (excretion within 24 h) and therefore, most likely reflect short-term or acute dietary intake [44, 48]. Long-term or habitual dietary intake might be best reflected by lipophilic compounds e.g. in blood or other less often investigated sample matrices such as hair, fat tissue or nail clippings [44, 48].

- **Robustness**:

  The term robustness addresses a number of factors that might confound the results and interpretation of a putative marker; it concerns among others the specificity of a marker. For the evaluation of robustness both controlled dietary interventions (controlled habitual diet; marker for one specific food, a food group or a whole dietary pattern [39, 46]), cross-sectional studies in free-living populations (suitability in different populations, interactions with other food or food matrices/background diet, amounts and frequency of consumption of such food) as well as metabolite profile databases of food components (occurrence and amounts of the marker in other food, variability) are necessary [44, 48, 72, 74]. The potential influence of other food components or the putative marker itself on the absorption, distribution, metabolism and excretion might be a relevant factor. However, these kinds of influences so far have been rarely investigated [74].

- **Reliability**:

  The reliability of a putative marker to indicate the true dietary intake should be confirmed by comparison with a gold standard or reference method for a different marker indicating dietary intake for the same food. These conformation studies are best carried out under highly controlled conditions. Another way to prove the reliability of a putative marker would be the comparison with traditional dietary assessment methods. However, this has to be done with care, as it will be unclear, whether a putative marker lead to different results than the traditional dietary assessment method, because of its unsuitability and unreliability (i.e. not specific for a certain food or not chemically or biologically stable), or whether it is due to the biased dietary assessment using self-reported intake [44, 74].

- **Stability**:
  Suitable protocols for sample collection, processing and storage should be developed and chemical stability (possible oxidation or pH-instability) and biological stability (metabolization via enzymes in the matrix) should be ensured by performing degradation tests [72, 74].
- **Analytical performance**:
  This criterion includes the overall validity of the analytical method, including but not limited to precision, accuracy, detection limits, inter- and intra-batch variability (repeatability), practicality [53, 72, 74]. The aim of this criterion is to assess whether a qualitative or quantitative analysis is feasible using the putative marker. Validation of the analytical method with a validated method, a reference method or by using reference materials is desirable.
- **Inter-laboratory reproducibility**:
  Results and data should be comparable, when the putative marker is measured by different laboratories [72, 74].

1.2.3 Identification of markers for health status exemplified by diabetes

In contrast to the subtle changes in metabolism through the diet, the onset and development of a disease may cause massive metabolic changes, detectable even though the metabolome is of course influenced by a multitude of factors, such as sex, age, diet, lifestyle, individual microbiota or drugs [53]. The growing number of chronical diseases such as obesity, diabetes, cardiovascular disease and in part cancer caused by metabolic dysregulations is nowadays a growing worldwide problem [75]. Diseases and inborn errors of metabolism cause characteristic alterations in metabolism, which are often apparent before clinical symptoms become evident [76]. The detection of early stages of diseases is of high importance, because this could enable more efficient treatments, prevention, a slower clinical progression and lower rates of mortality [76]. A whole number of diseases were investigated for early stage diagnosis such as cancer, diabetes, cardiovascular diseases or inborn errors of metabolism. However, not only early stage diagnosis is important, but also the classification and progression of the disease or prognosis of treatment success are highly relevant in clinical research [76]. Through the detection of several hundreds of metabolites at the same time, metabolomics and metabolite profiling approaches enable an efficient monitoring of perturbations in specific endogenous metabolites,

representing the complexity of key and potentially disease-related metabolic pathways [75-77]. Therefore, metabolomics can reveal crucial information related to the health status [77]. The comparison of metabolite profiles of perturbed and unperturbed metabolic pathways will enable deeper knowledge of underlying physiological and pathophysiological mechanisms of disease development and also identification of targets for treatment [75, 76]. Metabolomics has an immense potential for identification of biomarkers predicting the onset of diseases, their severity and progression, therefore enabling diagnosis, monitoring and therapy of diseases [15, 75-77]. Although the investigation of clinical biomarkers using metabolomics has been going on for a while, there is still a lot to be done, especially with respect to validation and application of these markers in a clinical setting, similarly as described above for nutritional markers [77].

Measurements of a static state in metabolism to determine health state has been the objective of many investigations, but in the last years, also the measurement of metabolism while challenging homeostasis has gained interest. In general, the metabolism is able to respond to internal and external stimuli in a dynamic way via a complex network of multiple mechanisms, pathways and interactions; thereby maintaining homeostasis [78]. Challenges such as dietary intake (acute exposition with e.g. high fat, high sugar, high caloric or high salt) or physical activity cause changes in metabolite concentrations that are related to physiological processes. A healthy organism is able to respond adequately to, and compensate these changes caused by stressors or other expositions by flexible metabolic adaptation via different feedback mechanisms, the buffer capacity of the homeostasis and, if necessary repair mechanisms [78-80]. A chronical exposure towards an unbalanced diet will cause a reduced, impaired or even a loss in flexibility of the metabolic responses and thereby cause the development of diseases such as diabetes, hypertension or cardiovascular disease [78-80]. Therefore, the flexible balancing reaction of the metabolism in response to an external challenge can determine the individual health status, the predisposition for diseases and enable an early diagnosis of many chronical diseases [40, 41, 78-80]. In the end, in comparison to measuring a static state, a challenge of homeostasis might be a more informative way for diagnosis via biomarkers, as it might enable a detection of more subtle and relevant effects that precede a disease, identify an individual predisposition towards certain diseases, or

predict the capacity of an organism to withstand stress [78, 79]. Untargeted metabolomics and metabolite profiling approaches are highly suited to monitor the time course of metabolites in body fluids after a challenge tests e.g. with an acute high intake of glucose, fructose, fat, salt or similar. An additional advantage and challenge at the same time is that the normal and healthy range of individual responses towards such challenges is broad, thereby enabling the classification of humans into metabotypes with similar phenotypic regulations in metabolism [40, 79, 81]. The key example of a challenge test is the traditionally used oral glucose tolerance test (OGTT) for the diagnosis of impaired glucose tolerance and diabetes. Here a patient receives a drink containing 75 g glucose and blood samples are taken before and during the test to determine how quickly glucose is cleared from the blood.

Diabetes mellitus and its related ensuing serious, life-threatening complications are one of the most significant global health problems, resulting in a large social, financial and health system burden across the world [82-85]. In year 2017, about 451 million adults (age 18-99 years) lived with diabetes worldwide, this number is expected to increase to 693 million by 2045 using a conservative estimation [86]. Almost half of all people living with diabetes are undiagnosed and moreover, an estimate of 374 million people lives with impaired glucose tolerance (IGT). Diabetes mellitus is a group of metabolic disorders characterized by prolonged high blood glucose levels as a result of either a relative or an absolute insulin deficiency. Insulin is the main anabolic hormone of the body, regulating among others, especially blood sugar level. It is produced by β-cells of the pancreatic islets or islets of Langerhans in the pancreas. Insulin promotes the absorption of glucose from the blood into liver, fat and skeletal muscle cells, where glucose is either used for energy production via glycolysis or converted to glycogen via glycogenesis or to fats via lipogenesis for storage. Therefore, blood glucose levels change even after food intake only within a narrow range (4,5-6,7 mmol/L). As a result of absolute or relative insulin deficiency, glucose is not absorbed properly into body cells, and therefore, prolonged high levels of blood glucose are observed (Table 1.2 for WHO definitions [87]). There are three main types of diabetes mellitus and pre-disease states [87]:

- **Pre-disease states**: IGT and impaired fasting glycaemia (IFG)
  Both are intermediate conditions between the healthy and diseased state, and are connected with a high risk for the development of type 2 diabetes mellitus.

IFG is defined by a consistently increased fasting blood glucose level, while IGT is defined by both increased fasting and 2-hour glucose.

- **Type 1 diabetes mellitus (T1DM)**: absolute insulin deficiency
  As a result of pathological alterations of the pancreatic islet β-cells, most often caused by autoimmune reactions in juvenile age, physiological insulin concentrations cannot be maintained. Symptoms such as excessive excretion of urine (polyuria), thirst (polydipsia), constant hunger, weight loss and fatigue may occur suddenly.
- **Type 2 diabetes mellitus (T2DM)**: relative insulin deficiency/insulin resistance
  Even though insulin concentration is mostly normal at the beginning the body is unable to effectively/properly respond to insulin. In later stages, a lack of insulin may also develop. Symptoms are similar to T1DM; however, they are less pronounced and progression is insidious. As a result, diagnosis often happens several years after the onset, when complications are already evident. T2DM was also called adult-onset diabetes, as it occurred in the past mainly in people over 40 years. However, recently it is occurring increasingly frequently also in younger people. The genetic predisposition, lifestyle factors such as a high fat, high sugar and high caloric diet, obesity and physical inactivity play a role in the development of T2DM.
- **Gestational diabetes**: occurring during pregnancy
  Gestational diabetes is marked by hyperglycaemia with blood glucose levels above normal but below those of diabetes. A consequence is an increased risk of complications during pregnancy and at delivery and additionally an increased risk for the women and their children to develop T2DM in the future.

Table 1.2. WHO diagnostic criteria and blood glucose levels [87].

Condition	Fasting glucose (mmol/L)	2-hour glucose (mmol/L)	HbA_{1C} (%)[1]
Healthy	<6.1	< 7.8	< 6.0
IFG	≥ 6.1 < 7.0	< 7.8	6.0-6.4
IGT	< 7.0	≥ 7.8	6.0-6.4
Diabetes mellitus	≥ 7.0	≥ 11.1	≥ 6.5

[1]glycated hemoglobin

90-95% of people with diabetes mellitus have T2DM [85]. Here, and also in the case of IGT and IFG, preventive measures such as lifestyle changes and especially weight reduction are viable options to ensure diabetes is never manifesting or even to cure T2DM during early stages. With the measurement of blood glucose, glucose in urine and glycated hemoglobin (HbA_{1C}) established indicators of diagnostic, progression and therapy monitoring are available [85]. Nonetheless, as mentioned above about half of the people with diabetes remain undiagnosed for years [86] and as a result develop severe complications such as diabetic retinopathy, diabetic nephropathy, diabetic neuropathy and diabetic angiopathy. Those long-term complications are mostly caused by the hyperglycaemia itself. Therefore, research for biomarkers of especially the prediabetic state and also biomarkers for the assessment of the risk to develop complications is of high importance. Metabolomics and metabolite profiling approaches may contribute to this field of work.

2 SCOPE AND AIM OF THIS THESIS

Research on markers for both dietary intake and health status using untargeted metabolomics and metabolite profiling approaches are emerging topics in nutrition and food sciences. Markers of dietary intake will in the future objectively complement the traditionally applied dietary assessment methods (such as FFQ or 24 h-dietary recall), which often lead to biased results due to intentional or unintentional misreporting. Furthermore, traditional clinical markers for diseases are most often indicators, when symptoms already are evident. New markers of health status or markers of disease will especially aim towards an early diagnosis of pre-disease states, where preventive measures are feasible options. Metabolomics with its ability to detect several hundreds of metabolites at the same time is a perfectly suitable tool for marker discovery. However, sometimes details within specific substance classes are overseen, due to the compromise between covering as many substance classes as possible at the same time and the necessity of dedicated sampling approaches for specific substance classes or groups. Therefore, dedicated metabolite profiling approaches might be a useful complement to untargeted metabolomics. Additionally, often the narrower perspective of metabolite profiling approaches will enable deeper elucidation and interpretation of a biological question within a specific metabolic pathway. Therefore, within this thesis, the application of dedicated GC-MS based metabolite profiling approaches (investigating volatile organic compounds and sugar compounds) for the discovery of markers for both dietary intake and health status was evaluated.

From everyday experience, it is known that urine smells differently after the ingestion of for example asparagus or coffee. And even with closed eyes, just smelling will enable humans to differentiate between lemons and strawberries. Although odorous volatile organic compounds are highly specific for foods, they have so far scarcely been targeted as potential markers of dietary intake. Coffee as a highly aromatic drink is a promising food to investigate the potential of volatiles as dietary markers. A HS-SPME-GC×GC-MS profiling method for volatile compounds was applied to identify markers of coffee consumption in urine samples of a cross-sectional study.

With respect to sugar compounds, so far attention was mostly paid to well-known sugar compounds, although numerous sugar compounds and isomers occur in nature, some of them highly specific for certain foods. Even in studies regarding

diabetes, a disease that is characterized by a disturbed carbohydrate metabolism, mainly glucose was of interest so far. Consequently, within this thesis, a dedicated GC-MS based metabolite profiling approach for sugar compounds was developed, optimized and applied towards identification of dietary markers within a cross-sectional study and also within an intervention study with an OGTT in healthy, prediabetic and diabetic volunteers.

3 PUBLICATIONS

3.1 Robust markers of coffee consumption identified amongst the volatile organic compounds in human urine

Carina I. Mack, Björn Egert, Erica Liberto, Christoph H. Weinert, Achim Bub, Ingrid Hoffmann, Carlo Bicchi, Sabine E. Kulling, Chiara Cordero (2019):

Robust markers of coffee consumption identified amongst the volatile organic compounds in human urine

Molecular Nutrition and Food Research 2019, 63, 1801060.

https://doi.org/10.1002/mnfr.201801060

RESEARCH ARTICLE

Nutrimetabolomics

Molecular Nutrition & Food Research
www.mnf-journal.com

Robust Markers of Coffee Consumption Identified Among the Volatile Organic Compounds in Human Urine

Carina I. Mack, Björn Egert, Erica Liberto, Christoph H. Weinert, Achim Bub, Ingrid Hoffmann, Carlo Bicchi, Sabine E. Kulling, and Chiara Cordero*

Scope: The human volatilome has gained high interest for the discovery of potential biomarkers of diseases. However, knowledge about the diet as a crucial factor affecting the volatilome is scarce. Therefore, the search for disease biomarkers, as well as the potential use of volatiles as dietary markers is so far limited. The aim of this study is to investigate the association of the diet with the urinary volatilome with the special task to find potential markers of coffee consumption in 24 h urine samples from the Karlsruhe Metabolomics and Nutrition (KarMeN) study.

Methods and results: Acidified urine samples are analyzed using an approach combining headspace solid phase microextraction (HS-SPME) sampling with untargeted GC×GC-MS. Overall, 138 reliably occurring volatiles are detected. To account for the unequally concentrated urine samples, results of six different commonly used normalization methods are compared. Statistical analysis evidences six potential markers of coffee consumption, the most promising being 3,4-dimethyl-2,5-furandione. A correlation analysis between the 24 h dietary recall data and the urinary volatilome reveals further promising associations.

Conclusion: The human urinary volatilome is highly affected by the diet, enabling access to a high level of information about potential diet-related biomarkers. Therefore, it is a very promising source for further investigations on dietary markers.

1. Introduction

Volatile organic compounds (VOCs) are formed by several chemical and biochemical processes and are therefore omnipresent in the environment.[1,2] VOCs play important roles in many different aspects of life such as aroma components of food, animal or plant hormones and pheromones, microbial products, combustion, or industrial processes. These exogenous volatiles can easily enter the human body via diverse paths (e.g., inhalation, ingestion, dermal absorption, and resorption in the gut), are distributed between different compartments of the body, and are excreted via diverse paths (e.g., exhalation, excretion, and secretion). At the same time the human body itself can produce a variety of VOCs endogenously through normal or abnormal physiologic processes or even through residing commensal microorganisms.[2–6] The combination of these exogenous and endogenous VOCs emitted by the human body are defined as the human volatilome, which can be measured in various specimen such as breath, skin emanations, sweat, saliva, blood, urine, or feces.[3,6,7] So far, the highly complex interplay between the various factors that influence the volatilome are not well understood.[4,5,8–11] Nonetheless, there is a growing interest to exploit the volatilome for early diagnosis of diseases[4,5,12] or metabolic disturbances such as, among others, cancer[9,13,14] and diabetes.[15,16] However, identifying distinguishable VOC profiles representing the healthy and diseased status is complicated by the huge interindividual variability caused by many factors such as genetic background, life style, diet, age, sex, physiologic differences in metabolism, or residing commensal bacteria.[4,5,8–11] Therefore, it is necessary to understand the interplay between and the impact of these influencing factors on the human volatilome.

Every-day experience demonstrates that diet may have a major (and sometimes olfactorily well-perceivable) impact on the urinary volatilome. So far, only a few systematic investigations of this phenomenon have been performed,[17–21] but it is likely that VOCs could be valuable for nutritional research. For example, with respect to epidemiologic studies where under- and/or overreporting complicate interpretation of results along the diet-health/disease-trajectory, valid dietary markers would be extremely useful.[22] While nonvolatile metabolites in urine as potential dietary markers have been in the focus of nutrition

C. I. Mack, B. Egert, Dr. C. H. Weinert, Prof. S. E. Kulling
Max Rubner-Institut
Department of Safety and Quality of Fruit and Vegetables
Haid-und-Neu-Straße 9, 76131 Karlsruhe, Germany
E-mail: sabine.kulling@mri.bund.de
Dr. E. Liberto, Prof. C. Bicchi, Prof. C. Cordero
Università degli Studi di Torino
Dipartimento di Scienza e tecnologia del Farmaco
Via Pietro Giuria 9, 10125 Torino, Italy
Prof. A. Bub
Max Rubner-Institut
Department of Physiology and Biochemistry of Nutrition
Haid-und-Neu-Straße 9, 76131 Karlsruhe, Germany
Prof. I. Hoffmann
Max Rubner-Institut
Department of Nutritional Behaviour
Haid-und-Neu-Straße 9, 76131 Karlsruhe, Germany

DOI: 10.1002/mnfr.201801060

research for some years, so far little attention has been paid to volatile metabolites.[23]

Breath and skin emanations are objects of several studies as they can be sampled easily and noninvasively.[5,7,9] Urine, as another readily accessible specimen, has gained popularity in the last years.[24,25] It has the additional advantage of preconcentration in the kidneys and of being more easily storable than, for example, breath or skin emanations. Even so, a huge disadvantage of urine is the differently diluted excretion, which necessitates (pre- or post-analytical) normalization. There are many normalization methods that can be applied, among others normalization on creatinine, osmolality, specific gravity, urine volume, mass spectral total useful signal (MSTUS),[26] or probabilistic quotient normalization (PQN).[27] So far, a scientific consensus on the evaluation and application of these methods is still missing.[28–31]

Coffee, with more than 800 volatile constituents or flavorants[17] and more than 150 million bags (weighing 60 kg) consumed worldwide in 2016,[32] is an important and promising food to study the informative potential of its volatiles as dietary markers. Wagenstaller et al.[17] already described an increase of some VOCs in urine after the ingestion of coffee in an acute intervention study with a targeted method based on a limited population ($n = 14$). Otherwise markers for coffee consumption are mainly from among nonvolatile metabolites.[23]

This study aims to identify potential markers of coffee consumption in 24 h urine samples from a selected subset of the observational Karlsruhe Metabolomics and Nutrition (KarMeN) study.[33] Urinary VOCs were sampled by headspace solid phase microextraction (HS-SPME) and analyzed by two-dimensional gas chromatography coupled with a quadrupole mass spectrometer (GC×GC–qMS). Different normalization methods on urinary VOC responses were tested to achieve a higher confidence level in results interpretation. Additionally, through the use of a 24 h dietary recall the association of diet with the urinary volatilome was investigated apart from coffee consumption.

2. Experimental Section

2.1. Study Design and Subjects

A subset of 24 h urine samples from the cross-sectional study KarMeN, performed at the Max Rubner-Institut in Karlsruhe, Germany (2011–2013), was analyzed. The study design and examination procedures were described elsewhere.[33] The study was approved by the ethics committee of the State Medical Chamber of Baden-Württemberg (F-2011-051) and was in accordance with the 1964 Helsinki declaration and its later amendments. The study was listed at the German Clinical Study Register (DRKS00004890) and has the WHO universal trial number: U1111-1141-7051. The subset consisted of 97 apparently healthy participants (18–80 years), of these 53 were male and 44 female. According to the conducted 24 h dietary recall, 48 participants drank coffee in the 24 h of urine collection (coffee consumers) and 49 participants did not drink coffee during that period (non-consumers). The subset was chosen randomly among coffee consumers and non-consumers of the KarMeN population with the precondition that about half of the coffee consumers and non-consumers, respectively, were male and the other half female. Additionally, for each, consumers and non-consumers, a uniform distribution of age was intended. For participant characteristics of the subset see **Table 1**.

Participants collected a 24 h urine sample in urine containers that were kept cooled throughout the collection period. The urine volume was recorded. A urine portion was centrifuged at 1850 × *g* at 20 °C and aliquoted. Aliquots were frozen at –20 °C for 1 day, and afterward cryopreserved at –196 °C until analysis. A quality control (QC) sample was prepared by mixing fresh spot urine samples of four male and four female volunteers shortly before the measurement series, then it was aliquoted and frozen for a short period at –20 °C until analysis.

The food consumption (in g d^{-1}) of each participant during the 24 h of urine collection was assessed in a personal interview using the 24 h dietary recall method and the software EPIC-Soft,[34,35] now renamed into GloboDiet.

2.2. Volatile Profiling of Urine Samples Using HS-SPME–GC×GC–qMS

2.2.1. Sample Preparation and HS-SPME Sampling Conditions

A short screening with different fibers (polyacrylate (PA), polydimethylsiloxane (PDMS), polydimethylsiloxane/divinylbenzene (PDMS/DVB), and divinylbenzene/carboxen/polydimethylsiloxane (DVB/CAR/PDMS)) as well as comparing sample preparation either with β-glucuronidase (*Helix Pomatia*, ≥100 000 units mL^{-1}, pH: 4.5–5; Sigma–Aldrich) or with addition of a pH modifier, i.e., HCl, was performed, before setting the method sampling protocol (see Section 1.1.1, Supporting Information).

Table 1. Participant characteristics.

Characteristics of participants	Total	Non-consumers		Coffee consumers	
		Women ($n = 22$)	Men ($n = 27$)	Women ($n = 22$)	Men ($n = 26$)
Age (years), mean	49.2	45.9	50.3	50.6	49.6
Age (years), range	18.9–80.4	18.9–77.4	20.2–74.3	22.7–80.4	21.2–79.5
Body mass index (kg m^{-2}), mean	24.1	21.5	25.2	24.7	24.6
Coffee consumption (g d^{-1}), mean	255.9	0	0	506.4	526.4
Coffee consumption (g d^{-1}), range	0–1330	0	0	50–1330	110–1320

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

Molecular Nutrition & Food Research
www.mnf-journal.com

The solid phase microextraction (SPME) fiber chosen for urinary VOCs sampling was the DVB/CAR/PDMS with d_f 50/30 µm and 2 cm length. A set of six fibers, all from Supelco (Bellefonte, PA), was conditioned before use as recommended by the manufacturer and tested for response standardization.

For each 24 h urine sample, a headspace crimp top vial (22 mL) was filled with 5 mL of deionized water to enable an efficient heat transfer toward the inert glass insert of 6 mL volume used for urine sampling. The reduced headspace volume enabled achieving adequate sensitivity for effective VOCs monitoring. To the 6 mL insert, 0.16 ± 0.001 g of NaCl was added together with 800 µL of urine sample. A total of 200 µL of HCl (1 mol L^{-1}) was finally added to lower the pH, before capping the vial and mixing with a Vortex unit. QC sample aliquots and study samples were prepared in the same manner. QC samples were used for the correction of batch or drift effects. Each day, 11 study samples, three QC samples, and one blank sample (consisting only of the internal standard solution) were prepared and analyzed. All study samples were measured twice.

For each run, 2 µL of internal standard solution (10 ppm tridecane in dibutylphthalate) was placed into a 22 mL glass vial for preloading onto the fiber for 30 min at 65 °C (standard-in-fiber procedure[36]), at the same time and temperature the urine sample was pre-equilibrated. Then, the fiber was exposed to the urine sample for 60 min at 65 °C. Afterwards, the fiber was introduced into the injector of the GC×GC system at 250 °C and thermally desorbed for 5 min.

2.2.2. Instrumental Setup for HS-SPME–GC×GC–qMS

HS-SPME–GC×GC–qMS analysis was performed on an Agilent 6890 unit coupled to an Agilent 5975C MS detector (Agilent, Little Falls, DE), equipped with an MPS-2 multipurpose sampler (Gerstel, Mülheim a. d. Ruhr, Germany) and a two-stage KT 2004 loop-type thermal modulator (Zoex Corporation, Houston, TX) cooled with liquid nitrogen. Data were acquired by Agilent MSD ChemStation ver. E.02.01.00. In the Table S1, Supporting Information, the different GC parameters including modulation parameters and the different MS parameters for the GC×GC–qMS measurement are shown.

2.2.3. Data Processing

Processing of raw data was performed using GC Image GC×GC Software version 2.5 (GC Image, LLC Lincoln NE). The sequential workflow described by Magagna et al.[37] was slightly modified and applied. After file import, rasterization, colorization, baseline correction, 2D-peak detection, and integration, a template with known analytes (targeted 2D peaks) was built using GC Image standard workflow.[37,38] Targeted peaks were identified by EI–MS fragmentation pattern (NIST MS Search algorithm, ver. 2.0, National Institute of Standards and Technology, Gaithersburg, MD; with Direct Matching threshold 900 and Reverse Matching threshold 950) with those collected in commercial (NIST2014 and Wiley7n) and in-house databases. As a further parameter to support identification, linear retention indices based on *n*-alkanes were considered and experimental values were compared with tabulated NIST reference indices. The target template consisting of 63 known volatiles was matched to all 2D-chromatograms and results (positive matches) saved (GC Project) to enable further processing.

Processed runs were then used to generate an untargeted/targeted (UT) template, following the UT fingerprinting principles,[37,38] including all reliable 2D peaks (known and unknown reliable peaks) that matched in all-but-one chromatograms of the set. The fully automated procedure is implemented in a dedicated software ambient named GC Investigator. The algorithm performs a comprehensive pair-wise peak matching and alignment of all 2D chromatograms based on retention times in both chromatographic dimensions and EI–MS fragmentation pattern (mode for preferred reliable peaks: most relaxed).

After application of the UT template to all samples (GC Project), the resulting 2D chromatograms were supervised and checked for the correct alignment of all reliable peaks. A preliminary data matrix with 172 features was obtained for which an analyte-specific batch correction was performed based on the measured QC samples. Afterward, analytes were evaluated with respect to their repeatability in QC samples (corrected interday coefficient of variance [CV] > 35% were excluded) and the occurrence of 2D peaks in heavily crowded chromatogram regions, leading to a final data matrix consisting of 138 peak features.

Before statistical analysis, missing values were replaced analyte-wise by half of the minimal value, which is a realistic value as they are mostly a result of the signal intensity being below the LOD. Afterward, normalization for the differential concentration of urine samples was carried out. For comparative reasons, not normalized data and data normalized using osmolality, creatinine, urine volume, MSTUS, and PQN were used. While scaling factors for osmolality, creatinine and urine volume are based on physiologic properties, those derived from MSTUS and PQN are data driven with the objective to prevent the impact of sample concentration variability. MSTUS is similar to normalizing on total signal, but uses only a subset of ions/features instead of including all measured ions (total ion count (TIC), which would also include noisy or specimen-specific ions/features) for a more reliable normalization.[26] PQN adjusts each sample of a metabolite profile according to the most probable dilution factor with regard to a reference sample to achieve a virtual overall concentration.[27] The reference sample could be either a selected biological study sample or a calculated median metabolite profile based on all measured biological samples, respectively. In recent publications,[31,39] when urine volume was used as the normalization reference, signal response was normalized by dividing by the urine volume and as this led to an increased biological variance, it was concluded that urine volume normalization is not appropriate.[31,39] However, compared to the direct proportionality between creatinine content or osmolality and the overall urine concentration, the urine volume is inversely correlated to the concentration of urine analytes (see **Figure 1**). Consequently, dividing by urine volume should lead to an increased variation contrary to the intention of normalization. A physiologically more reasonable option based on the principles of human renal physiology may be to normalize signal intensities by multiplication with urine volume. For comparison, both methods were applied in this manuscript (urine volume div and urine volume mult).

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

Molecular Nutrition & Food Research
www.mnf-journal.com

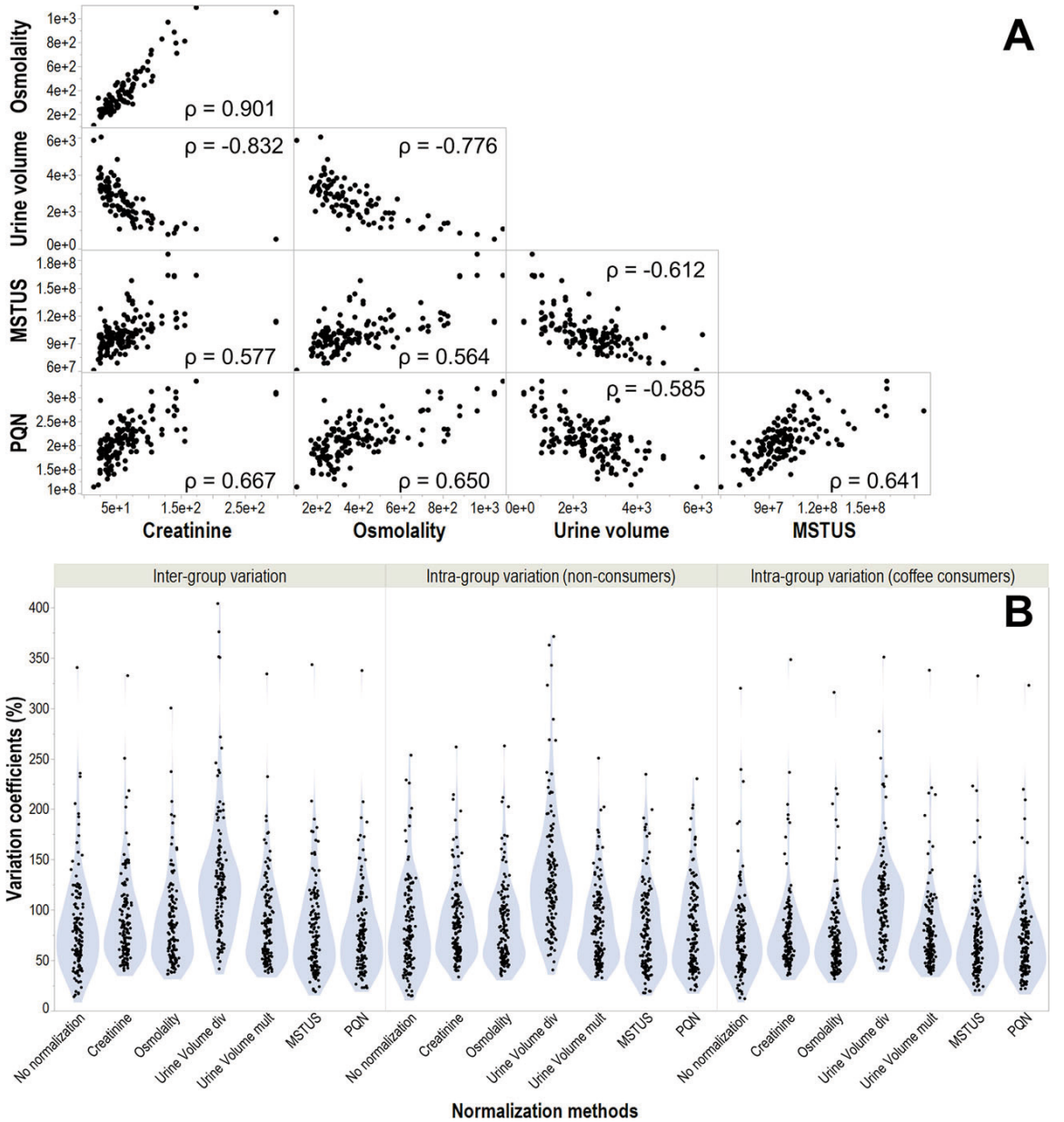


Figure 1. Normalization methods. A) Scatter plots of the different normalization methods including Spearman correlation coefficient. B) Violin plot for the CV's of all volatiles after the different normalization methods have been applied, grouped by inter- (CV over all study samples) and intra-group (CV within the two groups of coffee consumers and non-consumers) variation. The contour of the violins shows regions of data density, density is the highest where the violin is the broadest.

After assessment of data quality, some technical replicates were excluded. For the replicates, the mean was calculated and further statistical analysis and data interpretation were carried out on the resulting seven data matrices, one for each different normalization methods, as well as on the raw data with no normalization.

2.3. Statistical Analysis

Statistical analysis was carried out using JMP (version 13, SAS Institute Inc., Cary, NC, 1989-2007). Data matrices were built with 97 study samples (48 coffee consumers and 49 non-consumers) and 138 UT peak features.

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

Molecular Nutrition & Food Research
www.mnf-journal.com

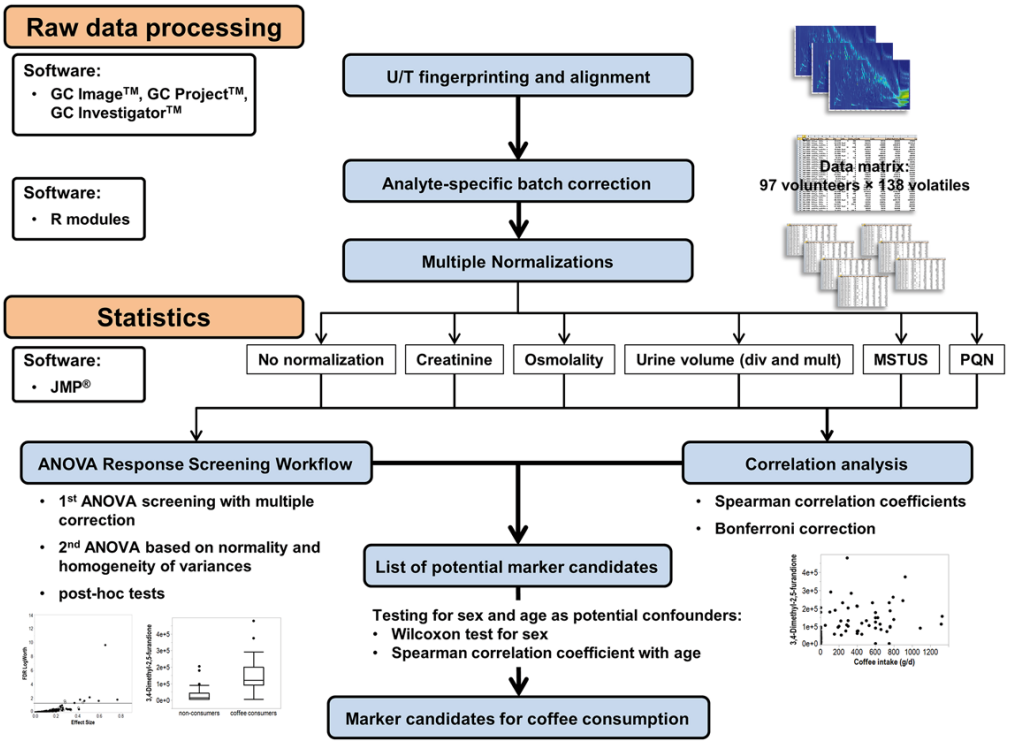


Figure 2. Workflow for the raw data processing and statistical analysis for potential coffee markers.

2.3.1. Comparison of Normalization Methods

In a first step, the overall impact of the different normalization methods on the data set was assessed using a correlation analysis. For this, Spearman correlation coefficients were calculated for the scaling factors used to normalize the data. In a second step, the CVs for all the 138 VOCs in all study samples and as well in the two groups of coffee consumers and non-consumers were calculated before and after application of all normalization methods.

2.3.2. Potential Markers of Coffee Consumption

First, an analysis of variance (ANOVA)-based response screening analysis including correction for multiple testing[40] was performed to identify volatiles differentiating the two groups of coffee consumers and non-consumers. The workflow included a first ANOVA with correction for multiple testing, a second ANOVA, and post hoc tests (each depending on the distribution and homogeneity of variance). The ANOVA based response screening was performed for the data without normalization, and the normalized data sets (creatinine, osmolality, urine volume div, urine volume mult, MSTUS, and PQN). Only significant volatiles showing a median fold change > 1.5 for at least one of the normalization methods were considered. Second, a correlation analysis using Spearman correlation coefficients was performed to consider the potential dose-response relation between the coffee consumption and the volatiles. Therefore, Spearman correlation coefficients of the volatiles with the amount of consumed coffee were calculated before and after normalization. Significance of the correlation was tested and only those reported that remained significant after Bonferroni correction, which in this case corresponded to $\rho > 0.35$. In **Figure 2**, a workflow for the raw data processing as well as the statistical analysis is depicted.

2.3.3. Relevance of Age or Sex for Potential Markers of Coffee Consumption

The possible influence of sex or age on the potential markers of coffee consumptions was evaluated. A Wilcoxon test was performed to ascertain the influence of sex, while Spearman correlation coefficients were calculated for the age effect.

2.3.4. Other Potential Volatile Markers Related to Food Consumption

To get a further impression on the usability of volatiles as potential markers for food consumption in general and the association of the diet with the urinary volatilome, the food consumption during the 24 h of urine collection (overall 81 different food consumption and nutrient intake variables, see ref.[41]) was correlated with the detected volatiles in 24 h urine samples before and after normalization. Only Spearman correlation coefficients with $\rho > +0.35$ for at least three normalization methods and with more than 25 participants who ate certain food items were

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

Molecular Nutrition & Food Research
www.mnf-journal.com

considered further. Occurring negative correlations of VOCs in urine with some food items were not considered at this stage because they could be related to endogenous metabolism and their relationship would be of a complex nature. Additionally, a Wilcoxon test was performed (lowest against highest tertile) for the normalization methods with the highest Spearman correlation coefficients.

3. Results

3.1. Analytical Characteristics

As a first step, different stationary phases of the fiber and sample treatment used for SPME were screened. Comprehensive approaches for selection of the fiber, sample treatment, and extraction conditions can be found elsewhere.[13,14,25,42–45] In **Table 2**, the number of detected 2D peaks is compared for the different fibers as well as the different sample preparations tested, including a blank sample with only glucuronidase and buffer. In Figures S1 and S2, Supporting Information, GC×GC–MS chromatograms for all fibers as well as the different sample preparations are depicted. The possibly gentler treatment with β-glucuronidase (*Helix pomatia*, ≥100 000 units mL^{-1}, pH: 4.5–5; Sigma-Aldrich) was excluded as the measurement of the blank enzyme sample revealed high abundance of VOCs (see Table 2 and Figure S2, Supporting Information). Enzyme preparations are usually isolated from living organisms (e.g., *Helix pomatia*) and can be complex mixtures. Therefore, it is highly recommendable to measure the blank enzyme before applying it to metabolomics studies.

Our short comparison gave the best result, i.e., the highest number of detected 2D peaks, for the DVB/CAR/PDMS fiber and the acidic treatment with HCl, which is in good agreement with literature evidences.[14,43,45] Therefore, these conditions were chosen for the analysis of the 24 h urine samples of the KarMeN study. Overall, 138 reliable 2D peaks were detected (at least in 70% of study samples) of which 63 were identified (targeted analytes). Repeatability of the measurement series was satisfactory with 66% of the volatiles having a CV < 25% in the batch corrected QC samples (see Table S2, Supporting Information).

3.2. Comparison of Normalization Methods

In Figure 1A, the scatter plots and Spearman correlation coefficients are shown for the different normalization methods. The associations between the normalization methods are of different type and strength. While normalization based on creatinine and osmolality is highly correlated ($\rho = 0.901$) and thus almost equivalent in this specific dataset, the effects of other normalization methods are obviously different and may lead to less comparable datasets and thus, at least in part, different results. Additionally, the inverse relationship of urine volume with the overall urine concentration and the other normalization methods can clearly be observed. In Figure 1B, the inter- and intra-group variation before and after the application of the different normalization methods is depicted in a violin plot. The densest region of the CV's was at similar values (CV ≈50%) before normalization and after normalization using creatinine, osmolality, MSTUS, and PQN. Interestingly, calculating total peak feature excretion per 24 h (urine volume mult) leads to a similar densest region, whereas normalization by dividing by urine volume (urine volume div) showed a broader distribution of CV's and the highest density lay at higher CV's (CV ≈125%).

Table 2. Comparison of detected 2D peaks for different fibers and different sample preparations.

	DVB/CAR/PDMS	PDMS/DVC	PDMS	PA
NaCl + HCl (1 mol L^{-1}) + urine	656	628	411	429
NaCl + buffer + β-glucuronidase + urine	640			
NaCl + buffer + β-glucuronidase + water	440			

3.3. Potential Markers of Coffee Consumption

The results of the ANOVA-based response screening (see **Figure 3**) and correlation analysis, as well as the potential influence of sex and age are summarized in **Table 3**. For the response screening, listed volatiles had a median fold change higher than 1.5 or, in case of the correlation analysis, a Spearman correlation coefficient above 0.35. In case of the ANOVA-based response screening, further volatiles were found to distinguish between coffee consumers and non-consumers, but these volatiles were increased in non-consumers and additionally showed only a weak median fold change (see Table S3, Supporting Information). Therefore, they are not discussed in this paper. The most interesting volatile marker candidate was 3,4-dimethyl-2,5-furandione, which was present in significantly higher levels in the samples of coffee consumers independent of the normalization method. This metabolite showed the highest fold changes (6.61–8.19) and correlations (ρ = 0.606–0.683). The other potential markers for coffee consumption were 2-methylfuran, guaiacol (i.e., 2-methoxyphenol), 2-/3-methylbutanoic acid, and 2-vinylfuran. The compounds exhibited a significant difference only for some of the different normalization methods (see Table 3). In case of methylbutanoic acid, both the 2-methyl- and 3-methylbutanoic acid isomer are coeluting, therefore, we report them as a sum parameter. To elucidate whether only 2-methyl-, only 3-methylbutanoic acid, or both are potential marker candidates, the raw data was reintegrated using a peak region over the coeluting isomers including two quantifier ions (m/z 57 and 60). This enabled a distinction of these two isomers. After applying the different normalization methods and performing a Wilcoxon-test as well as a correlation analysis with the coffee consumption, it becomes clear that both isomers were equally increased after coffee consumption (see Figure S3, Supporting Information).

In Figure S4, Supporting Information, both boxplots from the ANOVA screening as well as scatterplots from the correlation analysis are shown. 2-Methylfuran and 2-vinylfuran were found to significantly differ between males and females at least in one of the normalization methods, but not all. Median fold changes ranged between 1.18 and 1.38, therefore we believe the influence

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

Molecular Nutrition & Food Research
www.mnf-journal.com

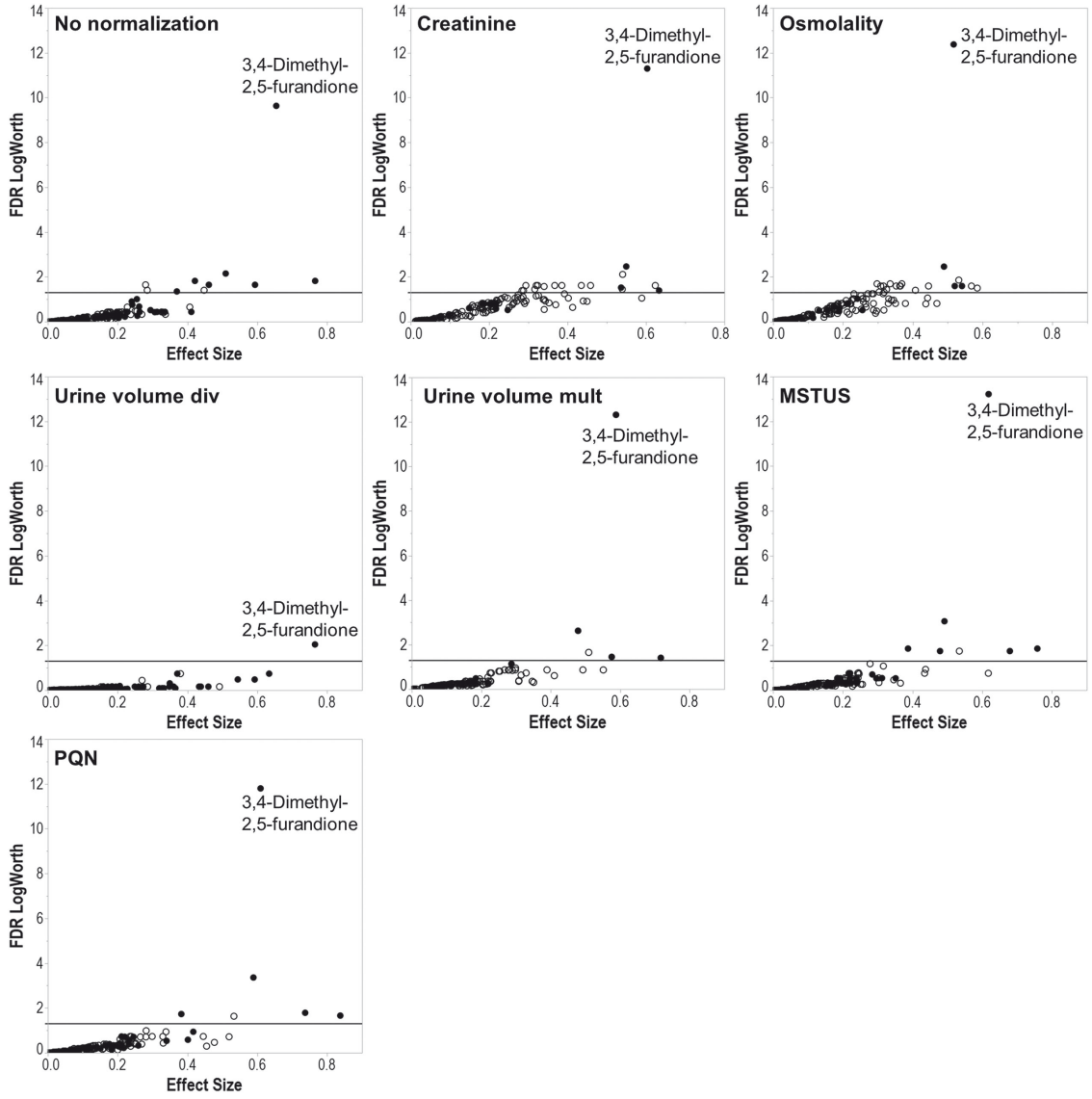


Figure 3. Results of the ANOVA based response screening for all normalization methods. FDRLogWorth = –log(*p*-value). Black line indicates *p*-value threshold of 0.05. ● Higher in coffee consumers; ○ higher in non-consumers.

of sex to be negligible. Guaiacol showed a rather weak correlation with age in case of two normalization methods.

3.4. Other Potential Volatile Markers Related to Food Consumption

With the aim to demonstrate the general capability of volatiles as potential markers of food consumption and to show the relevant associations of the diet with the urinary volatilome, we performed an additional Spearman correlation analysis with the food variables reported in the 24 h dietary recall during the 24 h of urine collection with the volatiles detected in these samples. Although the 97 participants were chosen with the specific aim to find markers of coffee consumption (see Section 3.3), we were able to detect further associations between volatile compounds in urine and dietary intake (see **Table 4**).

4. Discussion

4.1. Analytical Characteristics

Using the optimized HS-SPME–GC×GC–qMS method, we detected 138 volatile compounds reliably and reproducibly in 97 (24 h) urine samples of the KarMeN study. The 63 identified

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

Molecular Nutrition & Food Research
www.mnf-journal.com

Table 3. Results for potential markers of coffee consumption. In the first part "ANOVA screening", the median fold changes for those volatiles that were tested as significantly differentiating between coffee consumers and non-consumers after the post hoc tests are listed. The second part "Correlation analysis" lists the significant Spearman correlation coefficients of volatiles with the amount of consumed coffee after Bonferroni correction. The third part "Influence of sex" lists the significant *p*-values for a Wilcoxon test with males and females. The fourth part "Influence of age" lists the significant correlations of volatiles with age. If no value is included it was not tested as significant for this volatile or normalization method.

		3,4-Dimethyl-2,5-furandione	2-Methyl-furan	Guaiacol	2-/3-Methyl-butanoic acid[a)]	2-Vinylfuran
ANOVA screening (median fold change)	No normalization	7.81	1.83	1.55[b)]	1.69	1.68
	Creatinine	7.08	1.35	–	–	1.46
	Osmolality	6.61	1.42	–	1.39	1.33
	Urine volume div	8.19	–	–	–	–
	Urine volume mult	7.31	1.49	–	–	1.52
	MSTUS	7.36	1.76	1.51[b)]	1.48	1.58
	PQN	8.15	1.68	–	1.66	1.51
Correlation analysis (correlation coeff. ρ)	No normalization	0.683	0.541	0.486	–	0.432
	Creatinine	0.657	0.431	–	–	–
	Osmolality	0.673	0.493	0.360	–	0.359
	Urine volume div	0.606	0.404	0.384	–	0.364
	Urine volume mult	0.686	0.526	0.430	–	0.401
	MSTUS	0.682	0.559	0.515	–	0.406
	PQN	0.682	0.557	0.524	–	0.449
Influence of sex (*p*-values)	No normalization	–	–	–	–	–
	Creatinine	–	0.0242	–	–	–
	Osmolality	–	–	–	–	–
	Urine volume div	–	–	–	–	–
	Urine volume mult	–	0.026	–	–	–
	MSTUS	–	–	–	–	0.0362
	PQN	–	–	–	–	–
Influence of age (*p*-values)	No normalization	–	–	–[c)]	–	–
	Creatinine	–	–	0.352	–	–
	Osmolality	–	–	–[c)]	–	–
	Urine volume div	–	–	–	–	–
	Urine volume mult	–	–	–[c)]	–	–
	MSTUS	–	–	–[c)]	–	–[3]
	PQN	–	–	0.338	–	–

[a)] 2- and 3-Methylbutanoic acid are coeluting compounds; [b)] Tukey HSD as post hoc test, all other volatiles and normalization methods used Steel–Dwass as post hoc test; [c)] Correlations were tested as significant, but $\rho < |0.3|$.

VOC include different substance classes, among others alcohols, aldehydes, ketones, carboxylic acids, furan derivatives, terpenoids, and phenols. It is noteworthy that the 24 h urine samples were cryopreserved at –196 °C for 3–4 years and stored for a short period (3 months) at –80 °C before they were measured. Although investigations on the influence of storage conditions on the volatilome of urine advice not to store samples longer than 6–9 months at –80 °C,[44,46] we observed a complex profile of VOCs in measured 24 h urine samples. Nonetheless, it is to be expected that very volatile compounds will successively be lost even at cryopreservation temperatures (–196 °C). However, this will occur for all urine samples in a similar manner and therefore, a relative comparison as performed in this study is still feasible. Even after 3–4 years of cryopreservation, semi- and less volatile compounds are well detectable.

4.2. Normalization

In literature, many comparisons about the advantages and disadvantages of normalization methods can be found.[31,39] The most often applied criteria for comparing normalization methods is the number of distinguishing metabolites between groups. If considered exclusively, this criterion is highly questionable as a higher number of potential biomarkers does not necessarily prove the validity of these potential markers. Therefore, objective methods for the evaluation of normalization methods are needed. Some publications proposed application of multiple evaluation criteria (among others reduction of intra-group variation or consistency of identified markers).[28,29] Additionally, experimental design, the dataset, and the underlying research question are important criteria for selecting a normalization method.[29] The proposed schemes[28,29] are sophisticated approaches, but they

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

Molecular Nutrition & Food Research
www.mnf-journal.com

Table 4. Further associations of volatile compounds with dietary intake. The number of participants having ingested a certain food variable (n_{ing}), the Spearman correlation coefficients for all normalization methods and the p-values of the additionally performed Wilcoxon test (lowest against highest tertile) using the in bold highlighted normalization method with the highest ρ are listed.

Food variable	Volatile compound	n_{ing}	No normalization	Creatinine	Osmolality	Urine volume div	Urine volume mult	MSTUS	PQN	Wilcoxon
Bread sum	2-Pentylfuran	93	**0.452**	–	–	0.370	–	0.439	0.387	<0.0001
Cereal and cereal products	2-Pentylfuran	65	**0.420**	–	–	0.372	–	0.383	0.379	<0.0001
Polysaccharides	2-Pentylfuran	97	0.436	–	–	–	–	**0.459**	0.420	<0.0001
Baked goods	Octanoic acid	47	–	0.421	**0.422**	–	–	0.378	–	0.0005
Dairy products sum	Octanoic acid	85	0.394	0.363	–	–	–	0.408	**0.432**	0.0010
Milk sum	Octanoic acid	70	0.370	–	–	–	–	0.379	**0.397**	0.0002
Fat sum	2-Pentylfuran	80	–	–	0.357	–	–	**0.375**	0.373	0.0002
	2(3H)-Furanone, 5-ethenyldihydro-5-methyl-	80	0.356	–	–	–	–	**0.416**	0.404	0.0002
Sausage products	*p*-Cymene	50	0.437	0.353	–	–	–	**0.467**	0.447	<0.0001
	Calamenene	50	0.379	–	–	–	0.380	0.386	**0.391**	0.0039
Non-alcoholic beverages sum	Benzaldehyde	97	0.370	0.431	**0.455**	–	–	0.405	0.397	<0.0001
	4-Methylbenzaldehyde	97	–	0.391	**0.414**	–	–	0.362	–	<0.0001
	Butylbenzoate	97	–	0.446	**0.541**	–	–	0.446	0.371	<0.0001
	(14)	97	–	0.397	**0.437**	–	0.408	–	–	<0.0001
	(162)	97	–	0.368	**0.405**	–	0.384	–	–	0.0003
	Butyrolactone derivative	97	–	0.415	0.485	–	**0.489**	–	–	<0.0001
	(236)	97	–	0.383	0.415	–	**0.425**	–	–	0.0003
	1-Dodecanol	97	–	0.381	**0.467**	–	0.440	–	–	<0.0001
	Phthalic acid derivative	97	–	0.423	**0.528**	–	0.503	–	–	<0.0001
	Phthalic acid derivative	97	–	0.424	**0.501**	–	0.482	–	–	<0.0001
	Di-tert-butyl-phenol	97	–	0.418	0.456	–	**0.457**	–	–	<0.0001
	Dibutylmaleate	97	–	0.394	**0.505**	–	0.426	–	–	<0.0001
Alcoholic beverages	Ethanol	31	**0.569**	0.526	0.525	0.544	0.565	0.541	0.551	<0.0001
Carrot	(149)	29	0.379	0.459	**0.464**	–	0.412	0.393	0.407	0.0001
	(122)	29	–	**0.469**	0.458	–	0.375	0.369	0.378	0.0003
	(78)	29	0.354	**0.541**	0.527	–	0.389	0.417	0.436	<0.0001
	Methylsalicylate	29	–	0.420	**0.435**	–	–	0.355	0.367	0.0225
Fruit sum	4-Methylbenzaldehyde	78	0.369	0.390	**0.406**	–	0.380	–	0.359	0.0014
	(38)	78	–	0.4082	**0.4485**	–	0.4196	–	–	<.0001
	(40)	78	–	0.4179	**0.4565**	–	0.4087	–	–	<.0001
	(43)	78	–	0.3838	**0.4175**	–	0.3659	–	–	0.0002
	(68)	78	–	0.3889	**0.4048**	–	0.3634	–	–	0.0008
	Trimethyl-2-cyclohexenone	78	–	0.3703	**0.4052**	–	0.3539	–	–	0.0008
	(89)	78	–	**0.4147**	0.4073	–	0.3651	–	–	0.0001
	(98)	78	–	0.4378	**0.4773**	–	0.3981	–	–	<.0001
Apples raw	(24)	36	–	0.3953	**0.4161**	–	0.3799	–	–	<.0001
	(49)	36	–	0.3906	**0.4203**	–	0.3784	–	–	<.0001
	(58)	36	–	0.3636	**0.3964**	–	0.3874	–	–	<.0001
Nuts and seeds	(68)	25	–	**0.3557**	0.3537	–	0.3518	–	–	0.0008

necessitate a deep understanding of the implemented statistical and bioinformatical methods for their application. Additionally, if the implementation of these statistical and bioinformatical methods is not performed by oneself, complete transparency can hardly be guaranteed.

We applied six normalization methods (creatinine, osmolality, urine volume div, urine volume mult, MSTUS, and PQN), not with the aim of choosing one over another, but to objectively compare results with respect to their intrinsic similarity, their ability for reduction of inter- and intra-group variation and differences and similarities in the identified potential markers for coffee consumption (discussed in Section 4.3). For all applied normalization methods, correlations in different strength were observed (see Figure 1A). Therefore, statistical group comparison can be different depending on the applied normalization. All normalization methods as well as no applied normalization had

comparable variation (see Figure 1B), with the exception of normalization on urine volume div, which caused a higher inter- and intra-group variation. This is in agreement with other reports in recent literature[31,39] and confirms our assumption that normalizing on 24 h urine volume by division does not lead to an equalization of the differentially diluted urine samples, but to a higher variation due to its inversely proportional relationship with the overall urine concentration. Thus, the normalization on urine volume div cannot be recommended. Nevertheless, from the physiological point of view, urine volume can indeed be considered as a meaningful normalization reference for 24 h urine samples, because it allows the calculation of the total excretion of metabolites—which may, in case of an untargeted analysis and using the normalization on urine volume mult, be expressed as signal intensity per day, which is proportional to the excreted amount per day. This allows investigating associations of 24 h dietary intake and corresponding urinary metabolite excretion in the same 24 h. Twenty-four hour urine collections are not easily applicable in most studies and therefore, spot urine samples are taken instead, which necessitate normalization methods other than urine volume. On the other hand, as shown in Figure 1B, normalization using creatinine, osmolality, MSTUS, or PQN was obviously equally appropriate at the global level. Further, any normalization based on creatinine, osmolality, or urine volume is probably not appropriate if the urinary metabolite profiles of apparently healthy people and those with impaired renal function will be compared.

4.3. Potential Markers of Coffee Consumption

Using the ANOVA-based response screening and the correlation analysis, six potential markers for coffee consumption were found, 3,4-dimethyl-2,5-furandione, 2-methylfuran, guaiacol, 2- and 3-methylbutanoic acid, and 2-vinylfuran (see Table 3). Only 3,4-dimethyl-2,5-furandione was found to be significant with all normalization methods, while the other potential markers were only significant for some of the normalization methods. For the normalization on urine volume div, which even introduced variation to the data, only 3,4-dimethyl-2,5-furandione was found to be significant. Applying normalization can reduce overall variation and therefore enable marker identification, but unsuitable normalization can as well introduce additional and artificial variation and is therefore a potential source of errors. However, biologically highly relevant markers, as 3,4-dimethyl-2,5-furandione (see Figure 3), will be found independent of the normalization method. Therefore, comparing the influence of different normalization methods on a dataset is a worthwhile and useful step.

3,4-Dimethyl-2,5-furandione is the most promising of the found potential markers for coffee consumption (see Figure 3). It had by far the highest median fold changes and correlation coefficients, additionally neither sex nor age showed any influence on the urinary levels of this VOC. Most likely, 3,4-dimethyl-2,5-furandione is a product of the roasting process of coffee beans. It was described in the headspace of roasted Robusta coffee beans[47] and brews prepared with these beans,[48] but not in the headspace above green Robusta coffee beans.[47] Next to *Coffea canephora* (Robusta coffee), 3,4-dimethyl-2,5-furandione was as well detected in extracts from *Coffea arabica* varieties of different geographic origins.[49] Therefore, 3,4-dimethyl-2,5-furandione as potential marker of coffee consumption is plausible. However, roasting processes are applied to many foods and 3,4-dimethyl-2,5-furandione was also found in roasted chicory,[50] Ugandan vanilla beans (but not bourbon vanilla beans),[51] caramel,[52] and in flue-cured tobacco leaves,[53] affecting specificity of this potential marker candidate. Most of the mentioned foods are consumed rarely or in small amounts and the influence of smoking would have to be evaluated separately, because the KarMeN participants were all nonsmokers. Nonetheless, with high median fold changes (6.61–8.19), no overlap of the boxes in the boxplots of coffee consumers versus non-consumers (see Figure S4, Supporting Information) and good correlations ($\rho = 0.606$–0.683) in a cross-sectional study with participants on an unrestricted diet, 3,4-dimethyl-2,5-furandione remains a highly interesting and robust potential marker of coffee consumption. This is especially true because this marker was identified even though coffee is drunk in many different ways and recording details of coffee consumption are not part of a 24 h dietary recall. Not recorded details of coffee consumption are i) differentiated amounts of different coffee beverages (among others filtered coffee, instant coffee beverages or espresso), ii) brewing habit (e.g., amount and length), iii) degree of dilution with milk to some extent, iv) coffee variety (Arabica, Robusta, etc.), and v) the roasting process, its intensity, and the grounding. All these factors may be causal for the broad scattering of data points that can be observed in the scatter plots (see Figure S4, Supporting Information) and for this reason, correlation coefficients can only be indicators of associations in this case. An assessment and validation of the quantitative use of this marker is not feasible in this study, due to the unrestricted diet and the unknown details of coffee consumption. The quantitative application should be evaluated in a separate intervention study controlling these potential confounding factors. So far, 3,4-dimethyl-2,5-furandione seems to allow a robust, qualitative assertion about general coffee consumption.

The other significantly different VOCs (2-methylfuran, guaiacol, 2-methyl- and 3-methylbutanoic acid, and 2-vinylfuran) between coffee consumers and non-consumers had weaker effects (median fold changes < 1.9, $\rho < 0.56$). The influence of sex and age on these potential markers seems to be of negligible importance, as only in some of the normalization methods slight tendencies for differences in sex or age were observed (see Table 3). 2-Methylfuran, guaiacol, 3-methylbutanoic acid, and 2-vinylfuran were detected in roasted coffee and are—such as 3,4-dimethyl-2,5-furandion—a result of coffee roasting,[47,49,54] therefore they can be viewed as plausible markers. Specificity is similarly questionable due to their formation during roasting processes. 2-Methylfuran, 2-vinylfuran, and 2-methyl- and 3-methylbutanoic acid are formed during heating of glucose,[55] while guaiacol is formed from the thermal or microbial degradation of lignin or phenolic acids such as ferulic acid.[56] Once ingested with the coffee, potential volatile markers will enter the blood stream similarly to other food components.[12] From there, they are transported to various organs for elimination (e.g., to the lungs, where some are exhaled with the breath).[12] Others, such as the found potential markers of coffee consumption are removed in the kidneys and excreted with urine. Before excretion via urine often biotransformation in the liver (phase I and II metabolism) occurs. In this study, however, an acidic hydrolysis

of urine samples was performed to increase the number of observed VOC's, therefore, no assertion about potential phase II conjugation is possible.

Wagenstaller et al.[17] described guaiacol and 3-methylbutanoic acid to increase after coffee consumption in a targeted analysis of selected VOCs of urine samples obtained from 14 volunteers. 3,4-Dimethyl-2,5-furandion, our most relevant VOC, was not measured. Wagenstaller et al.[17] observed further VOCs to increase after coffee consumption, such as among others 4-vinylguaiacol and β-damascenone. These VOCs were also identified in our analysis, while others might have remained unidentified or were not detected. However, they were all below the significance threshold for marker candidates of coffee consumption in our statistical analysis. Most likely due to the more controlled conditions in an intervention study, the small number of volunteers and the measurement of spot urine samples in comparison to the KarMeN cross-sectional study with 97 participants on an unrestricted diet and the usage of 24 h urine samples, Wagenstaller et al.[17] were able to detect a larger number of potential marker candidates of coffee consumption. However, our identified markers may be considered as more robust against all kinds of potentially interfering factors (e.g., other food consumption) and are based on a broader population.

4.4. Other Potential Volatile Markers Related to Food Consumption

The aim of the correlation analysis between the 24 h dietary intake data and the VOCs in 24 h urine was to explore the potential of using VOCs as dietary markers, not a comprehensive evaluation of further potential dietary markers. In the literature, VOCs have gained major interest for the diagnosis of diseases and although many authors assume that the volatilome is influenced by our diet,[4,5,7,9,12] specific reports are still mostly missing. Our study shows that dietary intake has noticeable associations with the urinary volatilome. Even though this study was specifically designed to investigate the consumption of coffee and therefore the strongest associations were observed with coffee, we were nonetheless able to observe further interesting correlations (see Table 4). Next to these positive correlations some negative correlations ($\rho = -0.3514$ to -0.4522) were observable. Due to their more difficult interpretation with respect to food consumption and because most of the volatiles were not identified, they are not discussed here. Some of the positive correlations seem plausible, exemplary mentioned here: i) 2-pentylfuran is a known constituent of the volatile profile of bread;[57] ii) octanoic acid is a known constituent of milk and other dairy products; iii) *p*-cymene and calamenene are typical constituents of different culinary herbs applied for taste and as antimicrobial agent in sausages;[58–61] or iv) ethanol is part of alcoholic beverages. All these plausible dietary marker candidates highlight the future potential of investigations of the volatilome in this respect.

5. Concluding Remarks

Acidified urine samples of the KarMeN study showed a rich VOC profile with 138 reliable volatiles (untargeted and targeted analytes) within samples collected from coffee consumers and non-consumers. Storage for 3–4 years at –196 °C preserved the urinary volatilome information, enabling effective and consistent fingerprinting by HS-SPME–GC×GC–qMS, even without any particular storage containers designed with respect to preservation of VOCs. The urinary volatilome is suitable for the identification of dietary markers, which could be illustrated by six identified potential markers of coffee consumption, the most promising being 3,4-dimethyl-2,5-furandione. These potential markers are plausible, but due to their likely formation via roasting processes, their specificity is not ensured and necessitates future validation in further studies. Nonetheless, these markers were found in 24 h urine samples of 97 participants of an observational study on an unrestricted diet. The fact that this was possible, despite not recording details of coffee consumption in a 24 h dietary recall, underlines the robustness of the marker candidates. Applying five different normalization methods gave similar results. With the exception of normalization on urine volume by dividing by it (urine volume div), all other normalization methods showed similar variation in the data as no normalization method. Nevertheless, markers that are highly relevant and show huge biological differences, such as in case of 3,4-dimethyl-2,5-furandione, will be identified independent of normalization procedures. The additionally performed correlation analysis with the general food consumption during the 24 h of urine collection enabled us to illustrate the future potential of VOCs as dietary markers. Our data highlights the substantial impact of dietary intake on the urinary volatilome, therefore in future work, it is necessary to investigate these associations in more detail, also with respect to their influence on potential disease biomarkers.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

The KarMeN study was financially supported by the German Federal Ministry of Food and Agriculture (BMEL). The authors would like to thank the staff of the Human Nutrition Study Center at the Max Rubner-Institut for their excellent work in conducting the study. The authors would also like to thank Eva Hummel, Friederike Wittig, Yvonne Martiniak, and Kimberley Hoffmann for the dietary assessment as part of the KarMeN study. C.M. especially wishes to thank Chiara Cordero, Erica Liberto, and Carlo Bicchi for the opportunity to stay and work in their laboratory at the University of Turin, as well as her supervisor at the Max Rubner-Institut, Sabine Kulling, for supporting and enabling the stay. The work was part of the Foodball project (grant no. 2814ERA03E) funded by the German Federal Ministry of Food and Agriculture (BMEL). S.K., C.W., C.C., and C.B. designed the research project; S.K., A.B., and I.H. developed the KarMeN concept and design; C.M. conducted the analytical experiment; C.C. and E.L. supervised the analytical measurement; C.M., B.E., C.W., and C.C. analyzed data and performed statistical analysis; C.M. wrote the initial draft of the manuscript; C.W., S.K., C.C., and E.L. critically reviewed and contributed to the manuscript; C.M. and S.K. had primary responsibility for final content. All authors read and approved the final version of the manuscript.

Conflict of Interest

The authors declare no conflict of interest.

Keywords

coffee, dietary markers, nutrimetabolomics, untargeted HS-SPME-GC×GC-MS, volatilomics

Received: October 2, 2018
Revised: January 15, 2019
Published online:

[1] A. Zlatkis, R. S. Brazell, C. F. Poole, *Clin. Chem.* **1981**, *27*, 789.
[2] A. Buettner, M. Wagenstaller, J. Beauchamp, in *Flavour Development, Analysis and Perception in Food and Beverages* (Eds: Parker, J. K., Elmore, J. S., Methven, L.), Woodhead Publishing, Cambridge, UK, **2015**, Ch. 18.
[3] R. Cumeras, *Curr. Metabolomics* **2017**, *5*, 79.
[4] M. Shirasu, K. Touhara, *J. Biochem.* **2011**, *150*, 257.
[5] Y. Y. Broza, P. Mochalski, V. Ruzsanyi, A. Amann, H. Haick, *Angew. Chem., Int. Ed.* **2015**, *54*, 11036.
[6] A. Amann, B. de Lacy Costello, W. Miekisch, J. Schubert, B. Buszewski, J. Pleil, N. Ratcliffe, T. Risby, *J. Breath Res.* **2014**, *8*, 034001.
[7] B. de Lacy Costello, A. Amann, H. Al-Kateb, C. Flynn, W. Filipiak, T. Khalid, D. Osborne, N. M. Ratcliffe, *J. Breath Res.* **2014**, *8*, 014001.
[8] B. A. Kimball, *Bioanalysis* **2016**, *8*, 1987.
[9] K. Schmidt, I. Podmore, *J. Biomarkers* **2015**, *2015*, 1.
[10] M. Wagenstaller, A. Buettner, *Metabolomics* **2013**, *9*, 9.
[11] M. Wagenstaller, A. Buettner, *Metabolites* **2013**, *3*, 637.
[12] C. Angle, L. P. Waggoner, A. Ferrando, P. Haney, T. Passler, *Front. Vet. Sci.* **2016**, *3*, 47.
[13] C. L. Silva, M. Passos, J. S. Camara, *Br. J. Cancer* **2011**, *105*, 1894.
[14] M. Monteiro, M. Carvalho, R. Henrique, C. Jerónimo, N. Moreira, M. de Lourdes Bastos, P. G. de Pinho, *Eur. J. Cancer* **2014**, *50*, 1993.
[15] G. Rhodes, M. Miller, M. L. McConnell, M. Novotny, *Clin. Chem.* **1981**, *27*, 580.
[16] H. M. Liebich, O. Al-Babbili, A. Zlatkis, K. Kim, *Clin. Chem.* **1975**, *21*, 1294.
[17] M. Wagenstaller, A. Buettner, *Metabolomics* **2014**, *10*, 225.
[18] R. Roscher, H. Koch, M. Herderich, P. Schreier, W. Schwab, *Food Chem. Toxicol.* **1997**, *35*, 777.
[19] M. L. Pelchat, C. Bykowski, F. F. Duke, D. R. Reed, *Chem. Senses* **2011**, *36*, 9.
[20] A. Zeller, K. Horst, M. Rychlik, *Chem. Res. Toxicol.* **2009**, *22*, 1929.
[21] G. A. Zachariadis, A. V. Langioli, *Anal. Lett.* **2012**, *45*, 993.
[22] G. G. C. Kuhnle, *J. Sci. Food Agric.* **2012**, *92*, 1145.
[23] J. A. Rothwell, F. Madrid-Gambin, M. Garcia-Aloy, C. Andres-Lacueva, C. Logue, A. M. Gallagher, C. Mack, S. E. Kulling, Q. Gao, G. Praticò, L. O. Dragsted, A. J. G. Scalbert, *Genes Nutr.* **2018**, *13*, 15.
[24] S. M. Rocha, M. Caldeira, J. Carrola, M. Santos, N. Cruz, I. F. Duarte, *J. Chromatogr. A* **2012**, *1252*, 155.
[25] G. A. Mills, V. Walker, *J. Chromatogr. B Biomed. Sci. Appl.* **2001**, *753*, 259.
[26] Y. Gagnebin, D. Tonoli, P. Lescuyer, B. Ponte, S. de Seigneux, P. Y. Martin, J. Schappler, J. Boccard, S. Rudaz, *Anal. Chim. Acta* **2017**, *955*, 27.
[27] F. Dieterle, A. Ross, G. Schlotterbeck, H. Senn, *Anal. Chem.* **2006**, *78*, 4281.
[28] J. T. Bo Li, Qingxia Yang, Shuang Li, Xuehiao Cui, Yinghong Li, Yuzong Chen, Weiwei Xue, Xiaofeng Li, Feng Zhu, *Nucleic Acids Res.* **2017**, *45*, 162.
[29] A. M. De Livera, G. Olshansky, J. A. Simpson, D. J. Creek, *Metabolomics* **2018**, *14*, 54.
[30] A. M. De Livera, M. Sysi-Aho, L. Jacob, J. A. Gagnon-Bartsch, S. Castillo, J. A. Simpson, T. P. Speed, *Anal. Chem.* **2015**, *87*, 3606.
[31] B. M. Warrack, S. Hnatyshyn, K. H. Ott, M. D. Reily, M. Sanders, H. Zhang, D. M. Drexler, *J. Chromatogr. B* **2009**, *877*, 547.
[32] L. D'Elia, E. La Fata, F. Galletti, L. Scalfi, P. Strazzullo, *Eur. J. Nutr.* **2017**.
[33] A. Bub, A. Kriebel, C. Dörr, S. Bandt, M. Rist, A. Roth, E. Hummel, S. Kulling, I. Hoffmann, B. Watzl, *JMIR Res. Protoc.* **2016**, *5*, e146.
[34] N. Slimani, G. Deharveng, R. U. Charrondiere, A. L. van Kappel, M. C. Ocke, A. Welch, A. Lagiou, M. van Liere, A. Agudo, V. Pala, B. Brandstetter, C. Andren, C. Stripp, W. A. van Staveren, E. Riboli, *Computer Methods Programs Biomed.* **1999**, *58*, 251.
[35] N. Slimani, P. Ferrari, M. Ocke, A. Welch, H. Boeing, M. Liere, V. Pala, P. Amiano, A. Lagiou, I. Mattisson, C. Stripp, D. Engeset, R. Charrondiere, M. Buzzard, W. Staveren, E. Riboli, *Eur. J. Clin. Nutr.* **2000**, *54*, 900.
[36] Y. Wang, J. O'Reilly, Y. Chen, J. Pawliszyn, *J. Chromatogr. A* **2005**, *1072*, 13.
[37] F. Magagna, L. Valverde-Som, C. Ruiz-Samblas, L. Cuadros-Rodriguez, S. E. Reichenbach, C. Bicchi, C. Cordero, *Anal. Chim. Acta* **2016**, *936*, 245.
[38] D. Bressanello, E. Liberto, M. Collino, F. Chiazza, R. Mastrocola, S. E. Reichenbach, C. Bicchi, C. J. A. Cordero, *Anal. Bioanal. Chem.* **2018**, *410*, 2723.
[39] W. M. B. Edmands, P. Ferrari, A. Scalbert, *Anal. Chem.* **2014**, *86*, 10925.
[40] C. H. Weinert, M. T. Empl, R. Kruger, L. Frommherz, B. Egert, P. Steinberg, S. E. Kulling, *Mol. Nutr. Food Res.* **2017**, *61*, 1600651.
[41] C. I. Mack, C. H. Weinert, B. Egert, P. G. Ferrario, A. Bub, I. Hoffmann, B. Watzl, H. Daniel, S. E. Kulling, *Am. J. Clin. Nutr.* **2018**, *108*, 502.
[42] S. Zhang, L. Liu, D. Steffen, T. Ye, D. Raftery, *Metabolomics* **2012**, *8*, 323.
[43] B. Bojko, N. Reyes-Garcés, V. Bessonneau, K. Goryński, F. Mousavi, E. A. Souza Silva, J. Pawliszyn, *TrAC Trends Anal. Chem.* **2014**, *61*, 168.
[44] T. Živković Semren, I. Brčić Karačonji, T. Safner, N. Brajenović, B. Tariba Lovaković, A. Pizent, *Talanta* **2018**, *176*, 537.
[45] R. Cozzolino, L. De Magistris, P. Saggese, M. Stocchero, A. Martignetti, M. Di Stasio, A. Malorni, R. Marotta, F. Boscaino, L. Malorni, *Anal. Bioanal. Chem.* **2014**, *406*, 4649.
[46] S. Esfahani, N. M. Sagar, I. Kyrou, E. Mozdiak, N. O'Connell, C. Nwokolo, K. D. Bardhan, R. P. Arasaradnam, J. A. Covington, *Biosensors* **2016**, *6*, 4.
[47] E. Nebesny, G. Budryn, J. Kula, T. Majda, *Eur. Food Res. Technol.* **2007**, *225*, 9.
[48] G. Budryn, E. Nebesny, J. Kula, T. Majda, W. Krysiak, *Czech J. Food Sci.* **2011**, *29*, 151.
[49] M. W. Cheong, K. H. Tong, J. J. M. Ong, S. Q. Liu, P. Curran, B. Yu, *Food Res. Int.* **2013**, *51*, 388.
[50] H. H. Baek, K. R. Cadwallader, *J. Food Sci.* **1998**, *63*, 234.
[51] S. Zhang, C. Mueller, *J. Agric. Food Chem.* **2012**, *60*, 10433.
[52] L. Paravisini, A. Prot, C. Gouttefangeas, C. Moretton, H. Nigay, C. Dacremont, E. Guichard, *Food Chem.* **2015**, *167*, 281.
[53] Z. Xiang, K. Cai, G. Liang, S. Zhou, Y. Ge, J. Zhang, Z. Geng, *Anal. Methods* **2014**, *6*, 3300.
[54] M. S. S. Amaral, P. J. Marriott, H. R. Bizzo, C. M. Rezende, *Anal. Bioanal. Chem.* **2018**, *410*, 4615.
[55] I. S. Fagerson, *J. Agric. Food Chem.* **1969**, *17*, 747.
[56] R. Dorfner, T. Ferge, A. Kettrup, R. Zimmermann, C. Yeretzian, *J. Agric. Food Chem.* **2003**, *51*, 5768.
[57] A. N. Birch, M. A. Petersen, Å. S. Hansen, *LWT–Food Sci. Technol.* **2013**, *50*, 480.
[58] A. Meynier, E. Novelli, R. Chizzolini, E. Zanardi, G. Gandemer, *Meat Sci.* **1999**, *51*, 175.
[59] S. Barbut, D. B. Josephson, A. J. Maurer, *J. Food Sci.* **1985**, *50*, 1356.
[60] C. Busatta, R. S. Vidal, A. S. Popiolski, A. J. Mossi, C. Dariva, M. R. A. Rodrigues, F. C. Corazza, M. L. Corazza, J. Vladimir Oliveira, R. L. Cansian, *Food Microbiol.* **2008**, *25*, 207.
[61] M. Aminzare, J. Aliakbarlu, H. Tajik, *Vet. Res. Forum* **2015**, *6*, 31.

Supplemental Material

Robust markers of coffee consumption identified amongst the volatile organic compounds in human urine

Carina I. Mack, Björn Egert, Erica Liberto, Christoph H. Weinert, Achim Bub, Ingrid Hoffmann, Carlo Bicchi, Sabine E. Kulling, Chiara Cordero (2019):

Robust markers of coffee consumption identified amongst the volatile organic compounds in human urine

Molecular Nutrition and Food Research 2019, 63, 1801060.

https://doi.org/10.1002/mnfr.201801060

Supporting information for online publication

1. Materials and methods

1.1. Volatile profiling of urine samples using HS-SPME-GC×GC-qMS

1.1.1. Screening of fibres and comparison of sample preparation

The following four different fibre stationary materials were compared: polyacrylate (PA), polydimethylsiloxane (PDMS), polydimethylsiloxane/divinylbenzene (PDMS/DVB) and divinylbenzene/carboxen/polydimethylsiloxane (DVB/CAR/PDMS). For the comparison of different stationary phase materials, aliquots of the same urine sample were used and sample preparation and HS-SPME sampling conditions were as described for the final measurement series (see main paper section 2.2.1), meaning 0.16 g NaCl, 800 µL urine sample and 200 µL HCl (1 mol/L) with 30 min pre-equilibration at 65 °C, 1 h fibre exposure at 65 °C and 5 min desorption at 250 °C.

For the comparison of enzymatic hydrolysis with the acidic hydrolysis, the same urine sample as above was prepared with β-glucuronidase using the DVB/CAR/PDMS fibre for HS-SPME under the same conditions as described above. For enzymatic hydrolysis 0.16 g NaCl was weight out, 800 µL of urine, 800 µL of sodium acetate buffer (pH 5) and 20 µL of β-glucuronidase (*Helix Pomatia*, ≥100000 units/mL, pH: 4.5-5) were added and mixed. Afterwards, HS-SPME sampling was carried out as described above. To evaluate the potential influence of volatiles originating from the β-glucuronidase and the buffer, a blank sample using water instead of urine was also measured.

Supporting information for online publication

1.1.2. Instrumental setup for HS-SPME-GC×GC-qMS

Table S1. GC, modulation and MS parameters for measurement of volatiles using HS-SPME-GC×GC-qMS.

Parameter	Setting / value
GC parameters	
Carrier gas	Helium
GC mode	Constant flow
Initial column head pressure	276.4 kPa
Liner type	Splitless, straight, for SPME, Ultra Inert
^{1}D column	SolGel WAX; ^1L = 30 m plus; ^1d_c = 0.25 mm; ^1d_f = 0.25 µm
Modulator column	Inert column used in modulator and to transferline L = 1 m; d_c = 0.1 mm
^{2}D column	Mega OV1701; $^2L_{sep}$ = 1.0 m; ^2d_c = 0.10 mm, ^2d_f = 0.10 µm
Column connector	SilTite MicoUnion (SGE)
GC temperature ramp	40 °C (1 min) → 3 °C/min → 240 °C → 8 °C/min → 260 °C (9.83 min). Run time: 80 min.
Injection mode	split
Split ratio	1:5
Gas saver	2 min then 20 mL/min
Injection temperature	250 °C
Interface temperature	320 °C
Modulation parameters	
Modulator type	Cryogenic, liquid nitrogen-based, loop-type
Modulation period (P_M)	4 s
Hot jet temperature ramp	220 °C → 3 °C/min →320 °C
Hot jet duration	350 ms
MS parameters	
Ion source temperature	230 °C
Ionization mode	EI (70 eV)
MS Mode	Scan
Scan speed	12.500 amu/s
Scan range	0-45 min: m/z 40-240
	45 min-end: m/z 40-320
Data acquisition frequency	30 s^{-1}

Supporting information for online publication

2. Results

2.1. Analytical characteristics

2.1.1. Fibre and sample preparation comparison

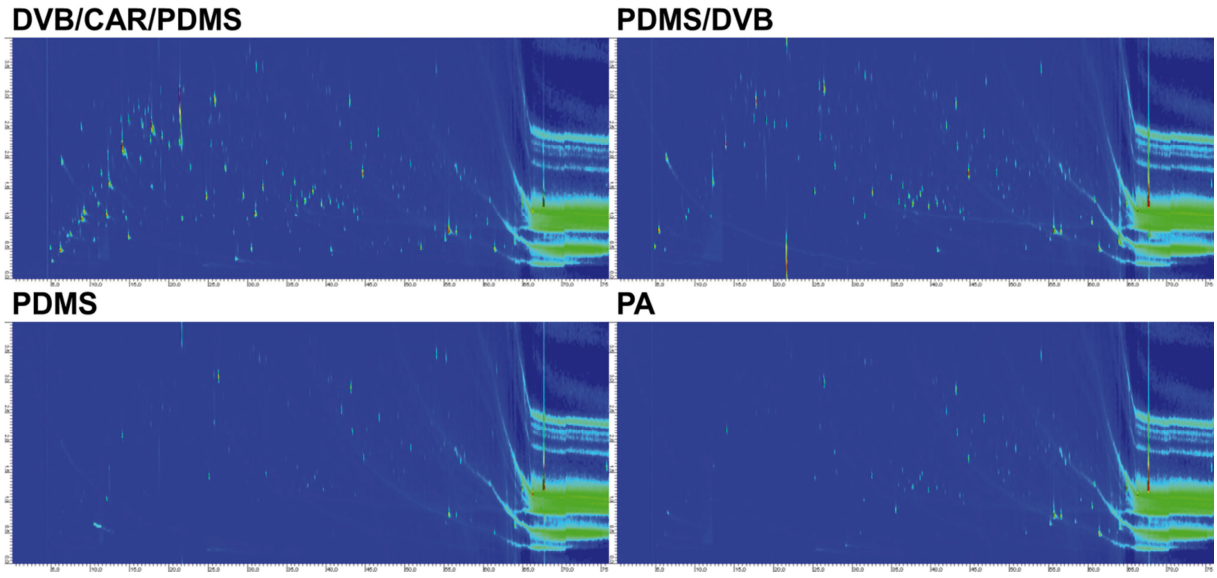


Figure S1. HS-SPME-GC×GC-MS chromatograms of a urine sample using different fibre stationary phases.

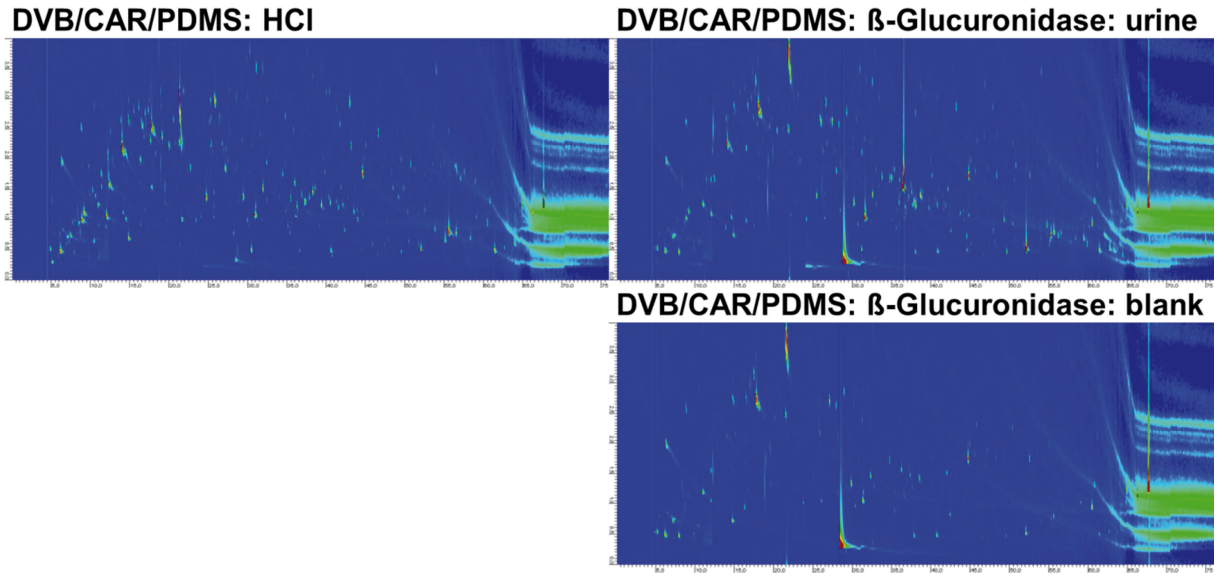


Figure S2. HS-SPME-GC×GC-MS chromatograms of a urine sample using different sample preparations.

Supporting information for online publication

2.1.2. Measurement serie – analytical characteristics

Table S2. Summary of all measured known and unknown volatiles, median peak volume and minimum to maximum range in study samples (not normalized), frequency in all study samples, and intra-day repeatability bevor and inter-day repeatability after batch correction in QC samples.

Volatile compound	Median peak volume	Minimum peak volume	Maximum peak volume	Frequency (%)	Intra-day$_{uncorr}$ (mean)	Inter-day$_{corr}$
(10)	1.10E+05	3.2E+04	7.3E+05	95.9	10.7	10.1
(11)	1.05E+05	2.3E+04	6.1E+05	97.9	22.8	23.2
3-Methyl-2-heptanone	1.68E+05	5.4E+04	1.0E+06	88.7	18.7	17.4
(14)	1.31E+05	3.7E+04	2.4E+05	100.0	26.0	23.5
4-Isopropylbenzaldehyde	8.94E+05	6.7E+04	1.0E+07	94.8	23.1	22.6
(24)	1.44E+05	2.5E+04	8.5E+05	95.9	15.8	19.5
(26)	2.44E+05	8.2E+04	9.4E+05	100.0	30.3	27.0
(30)	2.63E+05	1.0E+05	1.4E+06	100.0	24.5	24.0
(32)	1.85E+06	2.0E+05	1.2E+07	100.0	13.1	15.1
(33)	1.32E+06	1.3E+05	7.2E+06	100.0	9.7	10.6
(34)	3.32E+05	4.4E+04	3.3E+06	92.8	13.8	13.8
(35)	7.87E+05	5.1E+04	4.5E+06	100.0	15.2	16.0
(38)	1.68E+06	3.8E+05	6.9E+06	100.0	11.0	10.1
(39)	1.07E+06	5.4E+05	1.9E+07	100.0	23.5	20.2
(40)	1.06E+06	3.7E+05	4.2E+06	99.0	14.4	15.3
(41)	2.19E+06	6.7E+04	1.4E+07	80.4	8.2	7.0
(42)	9.48E+06	5.1E+06	1.2E+07	100.0	20.7	19.3
(43)	5.65E+06	1.8E+06	1.8E+07	100.0	16.5	15.0
(44)	3.63E+05	5.2E+04	4.1E+06	100.0	28.2	30.9
(49)	4.15E+05	5.8E+04	2.3E+06	100.0	9.6	10.0
(50)	1.08E+07	7.2E+06	1.6E+07	100.0	14.4	13.2
(51)	3.01E+05	1.2E+05	6.6E+05	92.8	15.6	16.7
(53)	8.77E+05	3.4E+04	4.3E+06	97.9	21.2	18.8
(54)	1.67E+05	5.0E+04	9.5E+05	97.9	28.4	24.4
(58)	5.09E+05	5.7E+04	2.5E+06	100.0	21.2	27.7
(65)	4.31E+05	2.7E+05	7.9E+05	100.0	16.6	15.1
(68)	5.00E+06	4.8E+05	2.0E+07	100.0	19.9	18.8
(69)	2.05E+05	6.0E+04	7.7E+05	91.8	15.2	14.7
(75)	2.30E+05	4.1E+04	1.3E+06	99.0	22.3	27.1
(78)	3.27E+05	8.7E+04	3.0E+06	99.0	12.1	13.1
(82)	1.03E+06	1.1E+05	6.1E+06	78.4	21.5	17.8
(83)	1.46E+05	5.1E+04	4.0E+05	100.0	34.3	32.1

Supporting information for online publication

Table S2 continued.

(84)	1.41E+06	1.0E+05	1.1E+07	99.0	16.3	16.7
(85)	7.92E+05	1.9E+05	2.2E+06	97.9	14.8	13.6
(86)	1.27E+05	4.7E+04	3.0E+05	94.8	11.3	11.2
(87)	2.09E+05	2.9E+04	1.2E+06	99.0	28.3	31.7
(88)	1.96E+05	6.4E+04	5.1E+05	99.0	20.0	19.8
(89)	3.21E+05	2.5E+04	4.8E+06	93.8	15.5	14.5
(90)	2.25E+05	3.0E+04	1.5E+06	94.8	15.4	14.3
(91)	2.01E+05	3.1E+04	9.2E+05	99.0	16.7	19.6
(92)	2.69E+05	6.4E+04	1.4E+06	94.8	24.9	21.6
(94)	3.91E+05	5.6E+04	3.1E+06	99.0	29.4	24.9
(95)	1.81E+05	5.8E+04	1.6E+06	97.9	37.3	30.9
(98)	1.39E+05	4.8E+04	3.3E+05	97.9	18.4	16.8
2-Vinylfuran	9.60E+04	3.2E+04	7.0E+05	94.8	35.7	30.7
(104)	2.41E+05	2.2E+04	2.5E+06	94.8	28.2	28.7
(118)	4.08E+05	5.0E+04	2.4E+06	100.0	14.6	18.5
(122)	3.08E+05	5.7E+04	3.4E+06	93.8	15.2	15.1
(128)	5.48E+05	8.6E+04	4.2E+06	96.9	14.1	13.3
(141)	1.44E+05	3.0E+04	9.2E+05	95.9	21.4	25.2
(142)	4.70E+05	7.4E+04	2.0E+06	97.9	23.1	22.1
(144)	8.98E+05	7.5E+04	3.3E+06	86.6	20.1	20.3
(149)	1.14E+05	3.3E+04	4.0E+06	79.4	23.6	19.7
(153)	1.27E+05	2.9E+04	6.0E+06	74.2	18.2	19.7
(159)	2.41E+05	3.8E+04	2.1E+06	89.7	11.9	11.5
(162)	1.90E+05	7.3E+04	6.8E+05	83.5	19.4	20.6
(174)	1.64E+05	3.4E+04	1.7E+06	80.4	31.7	27.5
(178)	2.10E+05	6.5E+04	1.5E+06	96.9	30.7	26.6
(183)	1.40E+05	3.3E+04	1.2E+06	70.1	25.3	22.6
(190)	2.83E+05	4.4E+04	1.5E+06	97.9	31.9	31.9
(197)	1.81E+05	3.1E+04	2.0E+06	88.7	30.8	28.3
(209)	1.56E+05	6.0E+04	3.3E+06	81.4	23.0	20.2
(212)	1.01E+05	2.8E+04	8.3E+05	72.2	31.9	32.5
(218)	1.53E+05	8.1E+04	3.2E+05	99.0	22.2	20.9
(220)	9.75E+04	3.1E+04	2.6E+05	100.0	32.7	32.5
(221)	3.53E+05	6.2E+04	6.1E+06	99.0	27.0	27.6
(223)	1.03E+05	4.2E+04	2.5E+05	100.0	31.4	26.6
(234)	1.90E+05	1.5E+04	1.5E+06	90.7	32.6	32.9
(236)	1.14E+05	5.3E+04	3.6E+05	99.0	20.6	18.6
4-Methylbenzaldehyde	6.45E+05	1.6E+05	1.9E+06	100.0	15.6	14.2
(253)	1.42E+05	4.0E+04	6.7E+05	88.7	16.8	15.6

Supporting information for online publication

Table S2 continued.

(254)	2.27E+05	5.9E+04	5.0E+06	100.0	27.6	26.9
(258)	1.38E+07	6.3E+05	8.6E+07	97.9	5.6	6.9
1-Dodecanol	2.92E+06	1.3E+06	8.2E+06	100.0	20.8	20.5
2(3*H*)-Furanone, dihydro-5-methyl-5-(2-methylpropyl)-	4.40E+05	9.8E+04	1.4E+06	100.0	11.5	11.8
2-Pentylfuran	2.01E+05	3.9E+04	1.1E+06	100.0	19.7	18.8
(36)	6.52E+05	1.9E+05	1.3E+06	100.0	23.5	26.2
(52)	2.80E+05	4.1E+04	1.4E+06	100.0	23.9	20.0
(268)	2.06E+05	4.5E+04	1.6E+06	97.9	28.9	28.5
(277)	9.09E+04	3.0E+04	3.0E+05	100.0	26.6	24.8
(278)	2.62E+05	4.8E+04	9.3E+05	100.0	22.8	25.7
1(3*H*)-isobenzofuranone (phthalide)	3.77E+06	1.8E+06	5.2E+06	100.0	17.1	14.8
1,2,3-Trimethylbenzene	3.05E+05	8.8E+04	1.0E+06	99.0	16.8	14.5
2-Ethyl-1-hexanol	9.26E+05	6.7E+05	4.0E+06	99.0	9.2	7.8
1-Terpineol	4.36E+05	3.6E+04	1.6E+07	99.0	14.9	12.7
1*H*-Indole	3.89E+05	3.7E+04	1.4E+06	100.0	30.1	30.2
2(3*H*)-Furanone, 5-ethenyldihydro-5-methyl-	5.41E+05	2.3E+05	1.9E+06	99.0	15.5	13.0
2(3*H*)-Furanone, 5-ethyldihydro-5-methyl-	7.19E+05	3.5E+04	1.0E+07	95.9	26.6	31.9
2,3-dehydro-α-ionone	1.63E+06	6.4E+05	1.0E+07	100.0	31.2	31.1
3,4-Dimethyl-2,5-furandione	9.62E+04	1.2E+04	4.8E+05	76.3	21.1	18.7
2-Acetylfuran	7.17E+05	2.1E+05	1.6E+06	100.0	17.1	15.0
2-Butanone	3.23E+05	7.9E+04	1.4E+06	100.0	31.1	31.1
3-Methyl-2-pentanone	9.34E+04	2.2E+04	8.6E+05	97.9	33.3	29.8
2-Thiophenecarboxaldehyde	8.49E+04	2.7E+04	4.2E+05	99.0	22.0	18.5
3-Furfural	4.79E+05	1.1E+05	4.1E+06	99.0	18.1	17.8
3-Hexanone	2.17E+05	6.0E+04	1.0E+06	100.0	32.7	30.3
4-Heptanone	9.17E+06	8.4E+05	5.1E+07	100.0	27.2	26.5
4-Vinylguaiacol	3.63E+05	7.6E+04	3.3E+06	94.8	33.7	30.1
5-Methylfurfural	1.05E+06	1.8E+05	4.1E+06	100.0	20.4	18.8
Benzaldehyde	1.16E+06	6.6E+04	3.5E+06	100.0	25.5	23.1
Benzene derivative	4.40E+05	1.1E+05	6.4E+06	99.0	19.6	16.1
Benzeneethanol	7.35E+04	4.0E+04	2.3E+05	87.6	15.6	13.3
Butylbenzoate	9.63E+06	4.0E+06	2.2E+07	100.0	17.9	21.7
cis-Theaspirane	3.82E+05	1.0E+05	8.0E+06	97.9	17.1	17.4
α-Pentylcinnamaldehyde	3.67E+05	5.6E+04	1.5E+06	100.0	31.9	32.2
Carvone	4.54E+05	2.7E+04	1.3E+07	88.7	39.3	33.0
Decanoic acid	1.84E+06	2.8E+04	4.3E+06	100.0	19.1	24.1

Supporting information for online publication

Table S2 continued.

Diethyl Phthalate	5.61E+05	2.8E+05	2.3E+06	100.0	34.6	33.9
Dihydromyrcenol	8.50E+05	8.8E+04	2.2E+07	95.9	17.7	15.8
Dimethyldisulfide	5.09E+06	1.4E+06	1.4E+07	100.0	24.6	26.1
2-Phenoxyethanol	1.78E+05	4.4E+04	4.0E+05	100.0	34.8	32.7
Furan	1.56E+05	4.3E+04	5.5E+05	100.0	28.8	25.5
2-Methylfuran	4.11E+05	9.0E+04	1.5E+06	100.0	34.5	30.6
Furan, tetrahydro-2,2-dimethyl-5-(1-methyl-1-propenyl)-	1.68E+06	1.3E+05	9.0E+06	82.5	29.5	29.4
γ-Terpinene	6.81E+05	5.8E+04	1.5E+07	81.4	16.9	16.3
Guaiacol	2.50E+06	5.1E+05	1.4E+07	100.0	10.5	10.0
2-Ethylhexanal	1.84E+05	4.9E+04	4.5E+05	96.9	21.9	21.2
Hexanoic acid	2.07E+05	1.1E+05	4.8E+05	99.0	14.8	13.8
Limonene	2.11E+05	4.6E+04	2.4E+06	100.0	28.7	27.5
Methanethiol	1.43E+06	5.5E+05	3.8E+06	100.0	29.9	26.0
Methylsalicylate	7.60E+05	1.1E+05	9.1E+06	100.0	28.2	29.6
Neoisomenthol	1.27E+05	2.9E+04	1.4E+07	81.4	14.4	16.4
Nonanoic acid	1.56E+06	4.5E+05	4.2E+06	100.0	24.8	21.8
Octanoic acid	3.50E+06	5.9E+05	1.2E+07	100.0	16.8	16.3
Phenol	3.88E+06	4.7E+04	3.0E+07	100.0	10.1	10.3
Piperitone	1.09E+06	4.7E+04	5.2E+07	93.8	23.7	28.1
trans-Linalool oxide	5.74E+05	8.8E+04	2.9E+06	94.8	23.8	20.6
trans-Theaspirane	3.28E+05	5.9E+04	2.9E+06	100.0	21.8	19.4
Thymol	1.04E+06	1.8E+05	1.5E+07	97.9	14.6	16.1
Dimethy-trisulfide	1.75E+06	1.6E+05	4.5E+06	100.0	24.1	25.4
beta-Damascenone	9.81E+05	2.0E+05	3.8E+06	100.0	31.2	25.3
2-/3-Methylbutanoic acid	5.97E+04	1.8E+04	2.8E+05	86.6	24.5	22.1
cis-Linalool oxide	2.85E+06	1.4E+05	1.1E+07	90.7	16.8	15.6
m-Cymen-8-ol	1.36E+06	1.3E+05	8.4E+06	97.9	22.7	20.1
p-Cymene	1.03E+07	1.1E+05	1.2E+08	99.0	21.0	23.3
p-Cymenene	1.97E+07	7.8E+05	8.9E+07	100.0	14.0	13.8
p-Methylguaiacol	2.83E+05	7.1E+04	3.6E+06	96.9	10.8	10.4
trans-Ocimenol	1.39E+05	3.5E+04	1.3E+06	69.1	40.5	33.6

Supporting information for online publication

2.2. Potential markers for coffee consumption

Table S3. List of volatile compounds testing as significantly higher in non-consumers than in coffee consumers after the post hoc test and their median fold changes. No value for the median fold change is included, if the volatile compound was not tested as significant for the respective normalization method.

Volatile compounds [a)]	**No normalization**	**Creatinine**	**Osmolality**	**Urine volume div**	**Urine volume mult**	**MSTUS**	**PQN**
(26)	-	-1.93	-1.92	-	-	-	-
(118)	-	-1.72	-1.90	-	-	-	-
4-Methylbenzalde-hyde	-1.71	-2.05	-1.89	-	-	-1.75	-
Benzaldehyde	-1.55	-2.22	-1.81	-	-	-1.65	-
(141)	-1.38	-1.58	-1.74	-	-1.40	-1.40	-1.47
trans-Linalooloxide	-	-	-1.70	-	-	-	-
(43)	-	-1.66	-1.66	-	-	-	-
Piperitone	-	-	-1.64	-	-	-	-
(277)	-	-1.30	-1.62	-	-	-	-
(35)	-	-1.63	-1.51	-	-	-	-
(86)	-	-1.57	-1.51	-	-	-	-
(52)	-	-1.50	-1.47	-	-	-	-
Methylsalicylate	-	-1.50	-1.35	-	-	-	-
(40)	-	-1.55	-1.34	-	-	-	-

[a)] Steel-Dwass as post hoc test for all volatiles listed here.

Supporting information for online publication

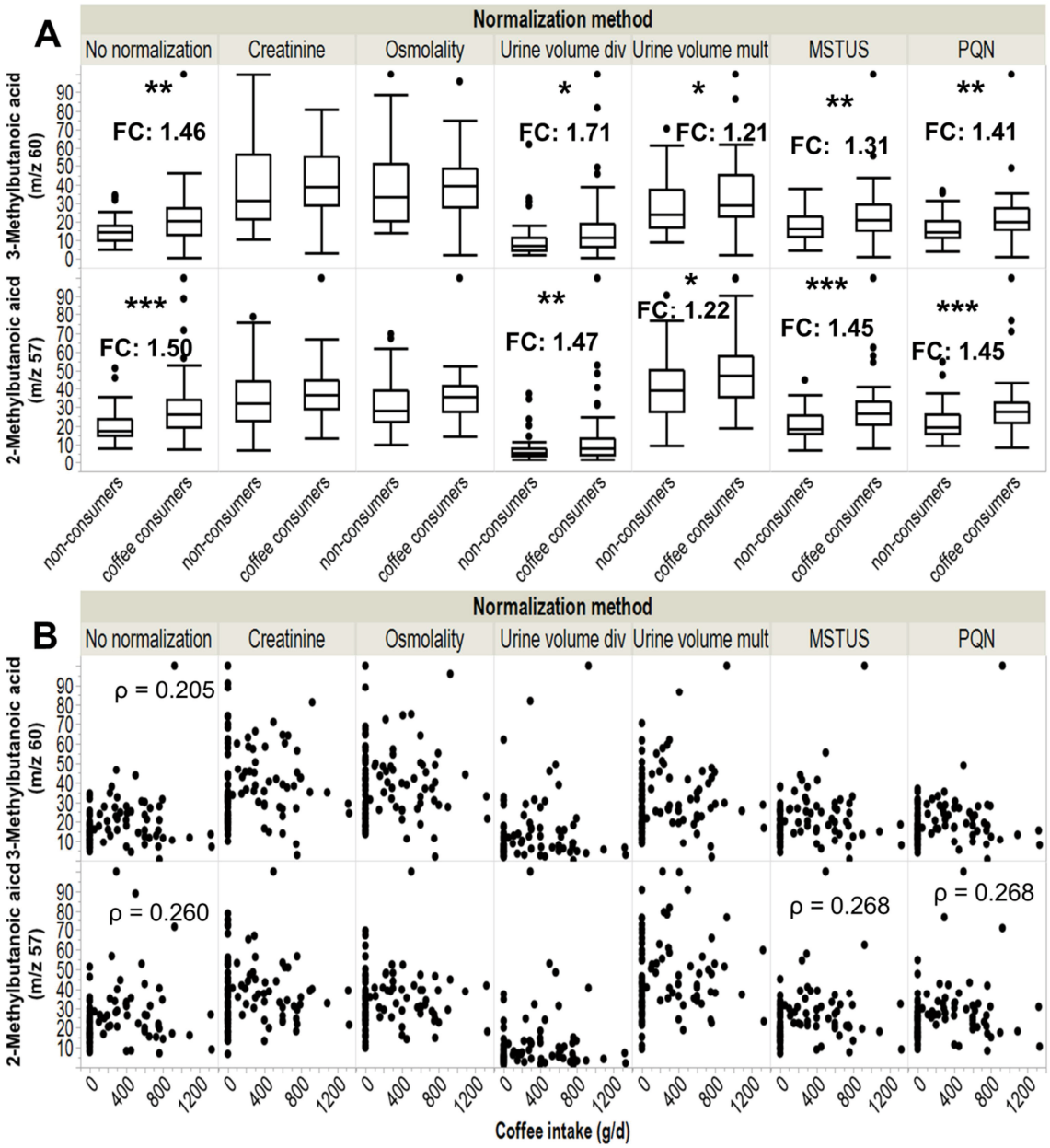


Figure S3. Separate isomers of methylbutanoic acid based on masses 57 and 60. Panel A: Boxplots of 2-methyl- and 3-methylbutanoic acid for coffee consumers and non-consumers for all normalization methods. *p < 0.05; **p < 0.01; ***p < 0.001; FC: median fold change. Panel B: Scatterplots of 2-methyl- and 3-methylbutanoic acid for coffee consumption (g/d) for all normalization methods. To achieve comparable axes for the different normalizations, signal intensities were scaled on the highest signal intensity to be 100%.

Supporting information for online publication

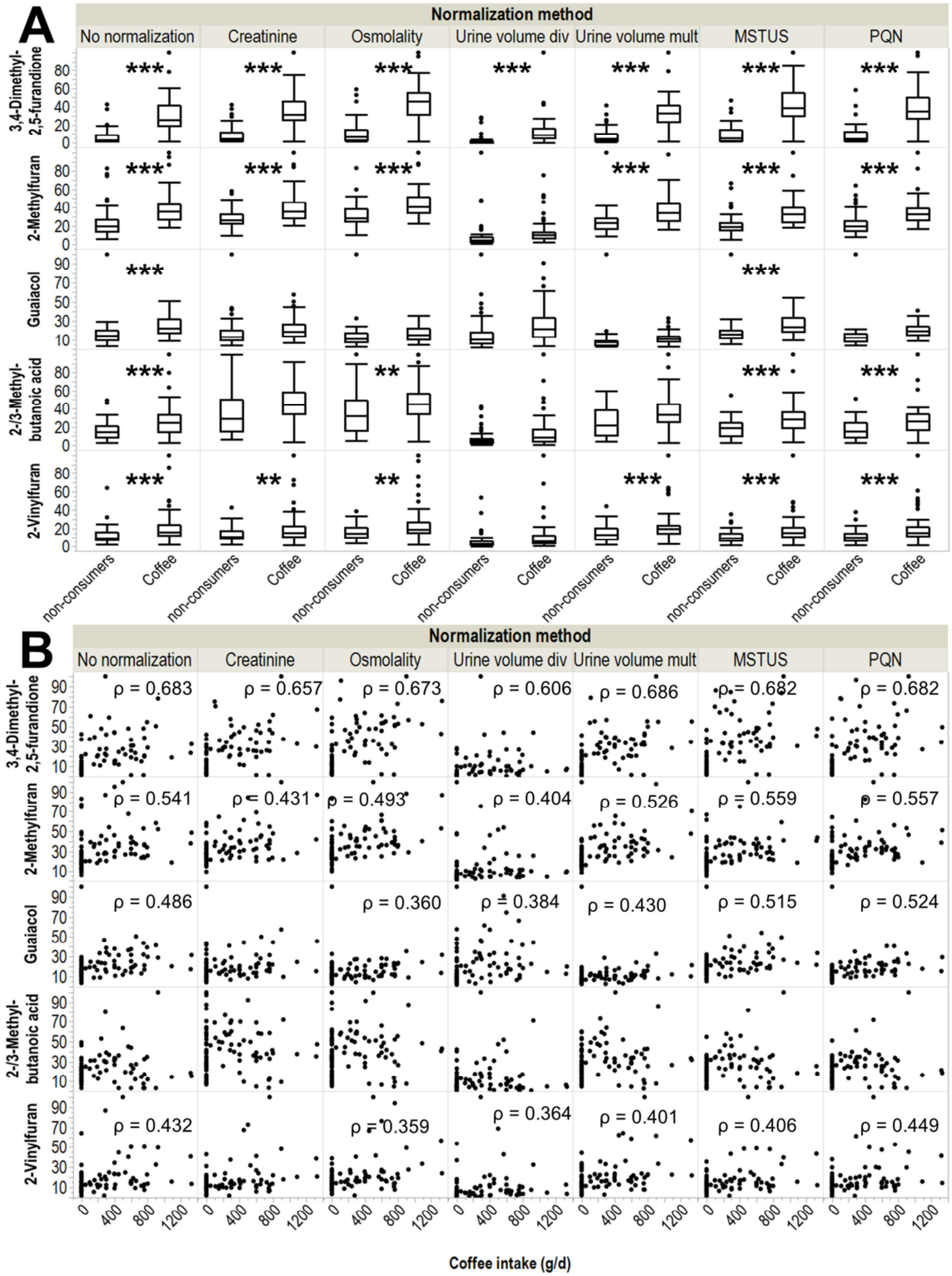


Figure S4. Potential markers of coffee consumption. Panel A: Boxplots of the potential volatile markers for coffee consumers and non-consumers for all normalization methods. *p < 0.05; **p < 0.01; ***p < 0.001. Panel B: Scatterplots of the potential volatile markers for coffee consumption (g/d) for all normalization methods. To achieve comparable axes for the different normalizations, signal intensities were scaled on the highest signal intensity to be 100%.

3.2 The complex human urinary sugar profile: determinants revealed in the cross-sectional KarMeN study

Carina I. Mack, Christoph H. Weinert, Björn Egert, Paola G. Ferrario, Achim Bub, Ingrid Hoffmann, Bernhard Watzl, Hannelore Daniel, Sabine E. Kulling (2018):

The complex human urinary sugar profile: determinants revealed in the cross-sectional KarMeN study

The American Journal of Clinical Nutrition 108(3): 502-516

https://doi.org/10.1093/ajcn/nqy131

See corresponding editorial on page 433.

The complex human urinary sugar profile: determinants revealed in the cross-sectional KarMeN study

Carina I Mack,[1] Christoph H Weinert,[1] Björn Egert,[1] Paola G Ferrario,[2] Achim Bub,[2] Ingrid Hoffmann,[3] Bernhard Watzl,[2] Hannelore Daniel,[4] and Sabine E Kulling[1]

Departments of [1]Safety and Quality of Fruit and Vegetables, [2]Physiology and Biochemistry of Nutrition, and [3]Nutritional Behaviour, Max Rubner-Institut, Karlsruhe, Germany and [4]Department of Food and Nutrition, Technical University of Munich, Freising-Weihenstephan, Germany

ABSTRACT

Background: Although sugars and sugar derivatives are an important class of metabolites involved in many physiologic processes, there is limited knowledge on their occurrence and pattern in biofluids.

Objective: Our aim was to obtain a comprehensive urinary sugar profile of healthy participants and to demonstrate the wide applicability and usefulness of this sugar profiling approach for nutritional as well as clinical studies.

Design: In the cross-sectional KarMeN study, the 24-h urine samples of 301 healthy participants on an unrestricted diet, assessed via a 24-h recall, were analyzed by a newly developed semitargeted gas chromatography–mass spectrometry (GC-MS) profiling method that enables the detection of known and unknown sugar compounds. Statistical analyses were performed with respect to associations of sex and diet with the urinary sugar profile.

Results: In total, 40 known and 15 unknown sugar compounds were detected in human urine, ranging from mono- and disaccharides, polyols, and sugar acids to currently unknown sugar-like compounds. A number of rarely analyzed sugars were found in urine samples. Maltose was found in statistically higher concentrations in the urine of women compared with men and was also associated with menopausal status. Further, a number of individual sugar compounds associated with the consumption of specific foods, such as avocado, or food groups, such as alcoholic beverages and dairy products, were identified.

Conclusions: We here provide data on the complex nature of the sugar profile in human urine, of which some compounds may have the potential to serve as dietary markers or early disease biomarkers. Thus, comprehensive urinary sugar profiling not only has the potential to increase our knowledge of host sugar metabolism, but can also reveal new dietary markers after consumption of individual food items, and may lead to the identification of early disease biomarkers in the future. The KarMeN study was registered at drks.de as DRKS00004890. *Am J Clin Nutr* 2018;108:502–516.

Keywords: urinary sugar profile, monosaccharide, disaccharide, polyol, sugar acid, GC-MS, dietary marker, sex, human urine, KarMeN study

INTRODUCTION

A variety of structurally different sugar compounds is present in the human body and even more so in our diet. We use the terms "sugar compounds or sugars" to refer to the following substance classes: mono- and disaccharides, as well as derived compounds thereof like polyols and sugar acids. Currently, sugar compounds are usually analyzed in urine samples with a focus on individual substance classes and, to date, most studies in this area have been performed with only a very limited number of volunteers (summarized in **Supplemental Table 1**). Combining the results from these studies revealed a quite complex urinary sugar profile consisting of many different sugar compounds. This is surprising because most recent studies have investigated the role of sugar compounds in human body fluids and focused mainly on common and well-known sugar compounds. Sugars and sugar derivatives in urine reflect the sugar compounds consumed within the diet as well as from endogenous sources. Of note is that absolute sugar concentrations in urine are very low because numerous sugars are efficiently reabsorbed in kidney tubular cells. Nevertheless, sugar compounds in human urine appear to be suitable dietary markers and, in the future, may even serve as early disease biomarkers, but knowledge on all this is highly limited.

In a few studies, specific sugar compounds were described as dietary markers for individual food items with examples such as *chiro*- and *scyllo*-inositol for citrus fruit in serum (1),

Supported by the German Federal Ministry of Food and Agriculture (BMEL). In addition, the work was part of the Foodball project (grant no. 2814ERA03E) funded by BMEL.

Supplemental Tables 1–4 and Supplemental Figures 1 and 2 are available from the "Supplementary data" link in the online posting of the article and from the same link in the online table of contents at https://academic.oup.com/ajcn/.

Address correspondence to SEK (e-mail: sabine.kulling@mri.bund.de).

Abbreviations used: CART, classification and regression tree; GC-MS, gas chromatography–mass spectrometry; KarMeN, Karlsruhe Metabolomics and Nutrition; QC, quality control.

Received November 22, 2017. Accepted for publication May 17, 2018.

First published online September 11, 2018; doi: https://doi.org/10.1093/ajcn/nqy131.

threonate and threitol for fruit and vegetables in serum (2, 3), and some C4 and C5 sugar compounds for broccoli consumption in urine (4). Other studies suggested that a combination of urinary fructose and sucrose may reflect total sugar intake (5–7), which is especially relevant in view of associations between sugar intake and negative health outcomes such as an increased risk for cardiovascular disease mortality (8, 9).

Because sugar compounds are involved in a variety of disease pathways, urinary sugars could also serve as biomarkers in the health-disease trajectory. Abnormal concentrations of urinary sugar compounds have been described for conditions such as diabetes mellitus (10–14), uremia (12, 15), invasive candidiasis (16), enzyme deficiencies like galactosemia (17, 18), pentosuria (19), or other inborn errors of metabolism (20–25). It thus seems reasonable to state that "there is more than glucose to look at" (26, 27).

With respect to the more commonly analyzed sugars, such as fructose and sucrose, there is a need to understand which factors determine their urinary excretion (28). Even less is known about the origin, metabolism, and functions of polyols (10, 23, 25). In a recent study, a new pathway for erythritol production from glucose was described and erythritol excretion was demonstrated to be associated with weight gain (29). This study highlights our current limited knowledge on sugars and sugar derivatives in the human body beyond glucose and fructose. To bridge this knowledge gap especially from the physiologic and pathophysiologic point of view, new analytical methods offering comprehensive detection of a wide range of major and many minor sugar compounds for nutritional and clinical research are thus urgently needed.

Here, we present a semitargeted gas chromatography–mass spectrometry (GC-MS) profiling method for the detection of >50 known and unknown sugar compounds in human urine and its application to 24-h urine samples derived from the observational KarMeN (Karlsruhe Metabolomics and Nutrition) study with 301 healthy participants (30).

METHODS

Study design and subjects

The cross-sectional KarMeN study was performed at the Max Rubner-Institut in Karlsruhe, Germany, between 2011 and 2013. Details on the study design and examination procedures were previously described by Bub et al. (30). Briefly, a total of 312 healthy participants aged between 18 and 80 y, who gave their written informed consent and were willing and able to perform all examinations, were recruited. Participants were excluded if they had a history of prevalent or chronic disease, were smokers, or took any medication, hormones, or dietary supplements. Women who were pregnant or breastfeeding were also excluded. Eleven participants who completed the study had to be excluded for other reasons, such as diseases requiring treatment, cardiac complications, voluntary dropout, cancer history, and acute cold with medication (30). Thus, a total of 301 participants remained for statistical analysis, 172 of whom were men and 129 were women. The local ethics committee approved the study and it was in accordance with the 1964 Helsinki declaration and its later amendments. The study was registered at the German Clinical Study Register (DRKS00004890) and has the WHO universal trial number U1111-1141-7051.

Participants were asked for a 24-h urine collection. Throughout the collection, bottles were kept in cool bags with cooling units. At the study center, the volume of the received 24-h urine samples was recorded, 2 × 14 mL were centrifuged at 1850 × *g* at 4°C for 10 min and then separated into aliquots. Samples were initially frozen at −20°C for 1 d and then cryopreserved at −196°C until analysis. A quality control (QC) sample was prepared by pooling 24-h urine samples from KarMeN participants. Osmolality was assessed via freezing-point depression of 24-h urine samples with the use of a micro-osmometer (Advanced Miro-Osmometer model 3MO, Advanced Instruments, Norwood, MA).

For the day of the 24-h urine sample collection, trained study personnel assessed the food consumption of each participant (in grams per day) in a personal interview through the use of a 24-h dietary recall with the software EPIC-Soft (developed by the International Agency for Research on Cancer (IARC) in Lyon) (31, 32), now renamed as GloboDiet. The amount of different foods consumed per day was assessed with the use of a picture booklet providing photographs of portion sizes for various foods as well as household measures and standard portions. For further analysis, the reported foods were summarized into food group variables (see **Supplemental Table 2**). Additionally, based on the German Nutrient Database "Bundeslebensmittelschlüssel" (BLS, version 3.02) (33), the total energy intake (in kcal per day) and intake of nutrients were calculated.

Semitargeted GC-MS sugar profiling

A Shimadzu GCMS QP2010 Ultra instrument was used in Scan-/SIM (selected ion monitoring)-mode to achieve high selectivity and sufficient sensitivity while at the same time being able to detect a priori unknown sugar compounds. Additionally, some abundant nonsugar compounds could be analyzed via this method. **Table 1** and **Supplemental Table 3** list all compounds that were detectable via this method, including the target and reference ions used for integration. The structural similarity of sugar compounds enables the usage of only a few selected masses for selective relative quantitation in the urine matrix (see Supplemental Table 3 and **Supplemental Figure 1**). Analytical details regarding chemicals, sample preparation, instrument, method, and data processing parameters were described by Rist et al. (34). Briefly, 24-h urine samples were diluted according to osmolality (60 mosmol/kg), 40 µL were evaporated and then derivatized via a 2-step procedure with 15 µL methoxylaminhydrochloride solution (20 mg/mL in pyridine; 30 min, 70°C, 1000 rpm) and 50 µL *N*-methyl-*N*-(trimethylsilyl)trifluoroacetamide + 1% trimethylchlorosilane reagent (1 h, 75°C) before analysis. To remove slight drift and offset effects, the raw signal intensities were corrected through the use of QC sample-based local linear regression functions (35).

Statistics

For all statistical analyses, the software JMP (version 13, SAS Institute Inc., Cary, NC, 1989–2007) was used.

Association of the urinary sugar profile with sex

A matrix with all known and unknown sugar compounds and the information on sex and age of the KarMeN participants was used to build a decision tree with the CART (classification and regression tree) algorithm. This approach was used to

TABLE 1
Summary of all sugar compounds that were included in the statistical analysis (columns 1–6) and additional analytes (columns 7–9)

Polyols	Monosaccharides[1]	Disaccharides[1]	Sugar acids[1]	Other	Unknown sugar-like	Amino acids	Organic acids	Others
meso-Erythritol	Xylose	Disaccharide U26	Erythronic acid	Levoglucosan	Unknown U03	Serine	Tartaric acid	Creatinine-enol
Threitol	Arabinose	Sucrose	Threonic acid	Ethyl-β-glucuronide	Unknown U04	Threonine	Isocitric acid	Unknown U11
Polyol U02	Ribose	Lactose	Sugar acid U01		Unknown U05	Cysteine	Hippuric acid	Unknown U16
Xylitol	Fucose	Maltose	Sugar acid U06		Unknown U12	Phenylalanine	Quinic acid	
Arabitol	Psicose	Disaccharide U27	Xylonic acid		Unknown U24	Lysine		
Ribitol	Fructose	Disaccharide U28	Ribonic acid		Unknown U25	Tyrosine		
1-Deoxysorbitol	Allose	Disaccharide U29	Sugar acid U09					
Fucitol	Galactose		Arabonic acid					
Mannitol	Glucose		Glucuronic acid					
Sorbitol	Mannose		Mannonic acid					
Galactitol	Mannoheptulose		Galactonic acid					
chiro-Inositol	Sedoheptulose		Gluconic acid					
scyllo-Inositol	Monosaccharide U21							
myo-Inositol								
Perseitol								

[1]Generally, 2 derivatives are formed for reducing sugar compounds during methoximation and silylation. For reasons of readability, the chemically exact denomination of compounds was deliberately omitted and only the first of 2 derivatives are listed.

uncover associations between the urinary sugar profile and sex. Advantages of the CART algorithm are its ability to cope with missing (not detected) values and its ability to handle categoric and numeric values in parallel. Not detected values (usually the results of signals below detection limits) were treated by the algorithm as a separate level of the variable. Concerning differences in the sugar profiles between men and women, the focus in this work was primarily on sugar compounds that were detected in <75% of the KarMeN participants, and thus, are potentially more sex-specific (in a qualitative sense). Age was included as an additional continuous variable after the first split, thus allowing the observation of associations between age and sugar compounds. Splitting was only allowed when $-\log_{10}$ (P values) (calculated by JMP) were significant after Bonferroni correction, meaning a $-\log_{10}$ (P value) >3.1206.

After CART analysis, the nonparametric Wilcoxon test was generally used to test for significant differences between men and women for the 2 most important metabolites as well as to distinguish between the maltose excretion of pre- and postmenopausal women.

Association of the urinary sugar profile with diet

To assess associations of diet with the human urinary sugar profile, an exploratory correlation analysis was performed with the use of the variables derived from the 24-h dietary recall (food and nutrient intake) with detected urinary sugar compounds (listed in Supplemental Tables 2 and 3, respectively). In a first step (selection of interesting correlations), Spearman rank correlation coefficients were determined by the pairwise method (threshold $\rho < -0.30$ or $\rho > 0.30$) and evaluated in conjunction with scatter plots. In a second step, participants were divided into groups based on consumption of certain food items for promising correlations. A Wilcoxon test was performed to ascertain significance for these groups. If <100 participants consumed a particular food or nutrient, an equally large group of nonconsumers was randomly selected. If >100 participants were consumers, tertiles were built and a Wilcoxon test for the first against the third tertile performed.

Sugar screening in plant materials from fruit and vegetables

To assess the plausibility and specificity of some of the potential dietary markers for food consumption, a screening of sugar compounds in a range of fruit and vegetable varieties was performed with the use of the same GC-MS profiling as for the urine samples. The aim was to screen as many fruit and vegetables as possible, but not to perform a comprehensive evaluation. Thus, only 1 pooled sample for each fruit and vegetable variety was measured.

Sample preparation for fruit and vegetables

Fruit and vegetables were bought from regional producers directly, weekly markets, or supermarkets. Overall, a total of 75 fruit and vegetable varieties (see **Table 2**) were purchased and, if possible, they were seasonally and regionally produced. The edible plant material of 5–20 fruits or vegetables (depending on fruit or vegetable size) was pooled into 1 sample, frozen in liquid nitrogen, and then coarsely ground and freeze dried for ≥3 d. The dried material was ground to a fine powder with a ball mill (Retsch MM400, Haan, Germany) for 20–60 s (depending on the consistence of the plant material) at 30 Hz and then stored at −80°C until analysis. For each sample, 20 ± 0.1 mg of freeze-dried powder was weighed out and then after addition of 20 μL of internal standard solution (pinitol and phenyl-β-glucopyranoside in water, each 2 mmol/L) extracted twice with 750 μL methanol for 10 min at 35°C and 1400 rpm. The collected supernatant was mixed and then centrifuged for 5 min at 4°C and 16,100 × g. After which, 20 μL of supernatant was evaporated and then derivatized with the use of the same 2-step procedure as for the urine samples, except that 40 μL of methoxylaminhydrochloride solution in pyridine (25 mg/mL) and 96 μL of *N*-methyl-*N*-(trimethylsilyl)trifluoroacetamide + 1% trimethylchlorosilane were used.

Semitargeted GCMS sugar profiling and data processing

The method for the measurement of the fruit and vegetable samples was the same as for the urine samples, except that the Rxi-5Sil-MS column was slightly shorter (54 m + 4 m precolumn), and as a result, time frames for SIM had to be adjusted (34). Each day, 30 fruit and vegetable samples, six 24-h urine QC samples, and a solvent blank were prepared and analyzed. Data processing was performed in the same way as for the 24-h urine samples (34).

RESULTS

Analytical performance of the semitargeted GC-MS sugar profiling method

Our newly developed and validated semitargeted GC-MS sugar profiling method (see **Supplemental Tables** 3 and **4** and **Supplemental Figures** 1 and **2**) enables the sensitive detection and relative quantification of 55 major and minor sugar compounds (see **Figure 1** and Table 1) encompassing mono- and disaccharides, polyols, sugar acids, and as yet not identified sugar compounds (see Supplemental Figure 1).

The assignment of the analyzed known and unknown sugar compounds into the different substance classes is shown in Figure 1. If desired, some amino and organic acids can also be analyzed with the method described here and this leads to a total number of 68 integrated analytes (see Table 1 and Supplemental Table 3). All sugar compounds detected via the semitargeted GC-MS method and their signal intensity ranges are listed in **Figure 2**.

Our method is also characterized by very good long-term repeatability and intermediate precision (see Supplemental Table 4 and Supplemental Figure 2) as proven by measurement series comprising overall 456 runs (312 study samples plus 144 QC samples). Thus, the method is suitable for long-term measurement series of human biofluids in large study cohorts.

Sugar profiling in participants of the KarMeN study

Sugar profile of human urine and biological variability

To determine metabolite-specific differences in the interindividual, i.e., biological variability, the CVs of the measured sugar compounds across all 301 KarMeN participants were

TABLE 2
Summary of a sugar screening in 75 fruit and vegetable varieties[1]

	Psicose	Mannoheptulose[2]	Perseitol[2]	Mannitol[3]	Galactose	Threitol	Xylose	Polyol U02
Eggplant	—	5	—	tr	Middle	—	Low	tr
Avocado	—	100	100	tr	tr	—	tr	tr
Leaf spinach	tr	13	—	tr	Low	—	tr	tr
Cauliflower	—	1	—	tr	High	—	Middle	tr
Common bean	tr	6	—	<0.1	Low	—	Low	Low
Broccoli	—	5	—	tr	Low	—	Low	Low
Iceberg lettuce	—	3	—	<0.1	Low	—	Low	Low
Peas	tr	2	—	<0.1	Low	—	tr	Low
Lamb's lettuce	Low	27	—	tr	Middle	—	Low	High
Cucumber	—	4	—	<0.1	Middle	—	tr	Low
Carrot	—	29	—	2.6	Low	—	tr	Low
Potato	—	5	—	<0.1	Low	—	—	—
Garlic	—	—	—	tr	tr	—	—	tr
Kohlrabi	—	tr	—	tr	Low	—	tr	tr
Garden lettuce	—	11	—	tr	Low	—	Low	Middle
Red cabbage	—	tr	—	—	Middle	—	tr	tr
Pointed cabbage	—	tr	—	—	Low	—	tr	tr
White cabbage	—	tr	—	—	Low	—	tr	tr
Pumpkin	—	tr	—	<0.1	Middle	tr	tr	Low
Leek	—	tr	—	—	Low	—	tr	tr
Striped lentil	—	—	—	tr	tr	—	—	tr
Black lentil	—	—	—	tr	tr	—	—	—
Lentil, "Perla"	—	—	—	tr	tr	—	—	—
Corn	—	—	—	tr	Low	tr	tr	—
Green bell pepper	—	6	—	<0.1	High	—	Low	tr
Red bell pepper	—	7	—	<0.1	High	—	Low	tr
Hot pepper	—	22	—	<0.1	High	—	Low	tr
Button mushroom	—	1	<1	100	tr	—	—	tr
Shiitake	—	—	<1	77.4	tr	tr	—	—
Small radish	—	2	—	<0.1	Low	—	Low	tr
Radish	—	—	—	tr	Low	—	tr	tr
Beetroot	—	—	—	—	Middle	—	tr	tr
Pointed pepper	—	12	—	<0.1	Middle	—	tr	—
Soy	—	—	—	<0.1	Low	—	tr	tr
Green asparagus	—	—	—	<0.1	Low	—	Low	tr
White asparagus	—	—	—	<0.1	Low	—	Low	tr
Grape tomato	—	6	—	<0.1	Low	—	Low	—
Tomato, "Matina"	—	11	—	<0.1	High	—	Low	tr
Tomato, "Resi"	—	7	—	<0.1	High	—	tr	tr
Zucchini	—	—	—	<0.1	High	—	Low	Low
Onion	—	tr	—	tr	Low	—	tr	—
Pineapple	—	—	—	—	Middle	—	tr	—
Apple	—	6	—	0.1	Low	—	High	tr
Apricot	—	4	—	<0.1	Low	—	Low	tr
Banana	—	—	—	tr	Low	—	—	—
Pear	—	5	—	<0.1	Low	tr	Low	tr
Blackberry	—	4	—	<0.1	Low	—	Low	—
Clementine	—	—	—	tr	Low	—	Low	—
Strawberry, "Asia"	—	4	—	—	Low	—	High	tr
Strawberry, "Elsanta"	—	2	—	tr	Low	—	High	tr
Pomegranate	—	7	—	18.1	Low	tr	tr	tr
Grapefruit	—	tr	—	tr	Low	—	tr	—
Blueberry	—	10	—	tr	Low	—	Low	—
Raspberry	—	4	—	tr	Low	—	Middle	—
Honeydew melon	—	tr	—	<0.1	High	tr	tr	Low
Red currants	—	6	—	tr	Low	—	Low	—
Black currants	—	14	—	<0.1	Low	—	Low	—
Sour cherry	—	7	—	<0.1	Low	tr	Low	tr
Sweet cherry	—	1	—	<0.1	Low	—	Low	tr
Kiwi fruit	—	5	—	<0.1	Middle	—	tr	—
Mango	—	4	—	tr	Low	—	tr	—
Small yellow plums	—	6	—	<0.1	Low	—	Low	—

(*Continued*)

TABLE 2 (*Continued*)

	Psicose	Mannoheptulose[2]	Perseitol[2]	Mannitol[3]	Galactose	Threitol	Xylose	Polyol U02
Nectarine	—	4	—	<0.1	Low	—	Low	—
Orange	—	tr	—	—	Low	—	Low	—
Papaya	—	tr	—	<0.1	Low	tr	Low	tr
Passion fruit	—	2	—	tr	Low	tr	tr	—
Yellow peach	—	5	—	<0.1	Low	—	Middle	tr
White peach	—	4	—	<0.1	Low	—	Low	tr
Physalis	—	7	—	<0.1	Low	—	tr	—
Gooseberry	—	3	—	—	Low	—	Low	—
Red table grapes	—	4	—	—	Low	—	tr	tr
White table grapes	—	5	—	—	Low	tr	tr	tr
Lemon	—	—	—	tr	Low	—	tr	—
Plum	—	5	—	<0.1	Low	—	Low	tr
Fig	—	4	—	tr	Middle	—	tr	tr

[1]High, fruit or vegetables with signal intensities in the highest tertile of the signal intensities for this analyte; Low, fruit or vegetables with signal intensities in the lowest tertile of the signal intensities for this analyte; Middle, fruit or vegetables with signal intensities in the middle tertile of the signal intensities for this analyte; tr, trace analytes (analytes below a signal intensity of 8000).

[2]Ratios of mannoheptulose and perseitol signal intensities for the fruit and vegetables compared with avocado signal intensity.

[3]Ratios of mannitol signal intensities for the fruit and vegetables compared with button mushroom signal intensity.

determined. Some sugar compounds were excreted with a narrow concentration range, for example glucuronic acid with a CV of 29.8%, whereas others showed a huge biological variability, such as lactose with a CV of 294.5% (see Figure 2). In addition, the relative frequency of occurrence of individual sugar compounds in the 24-h urine samples of KarMeN participants is listed in Supplemental Table 3. To further assess factors underlying the huge biological variability, analyses focused on sex as a determinant and on dietary intake reconstructed from dietary intake measures.

Association of the urinary sugar profile with sex

To identify sugar compounds associated with sex, a decision tree using the CART algorithm was built (see **Figure 3**). In

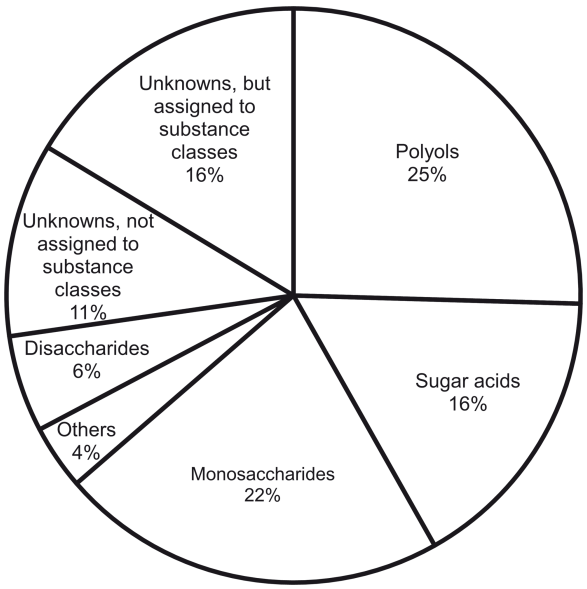


FIGURE 1 Classification assignment of sugar compounds into different substance classes.

Table 3, possible candidates for a split are given for the first 3 nodes and for leaves. Evidently, the most relevant sugar compound separating men and women was maltose, which was detected in 78.3% of women but only in 35.5% of men. In addition, the urinary maltose concentration was significantly higher in women ($P < 0.0001$, see Figure 3). Other important metabolites were gluconic acid, fructose, and an unknown sugar compound, which were found in >75% of the study samples and recently discussed by Rist et al. (34). In the second node, where age was included as an additional potential splitting candidate, the 3 top determinants for separating men and women were age, gluconic acid, and sedoheptulose. Interestingly, splitting on the basis of sedoheptulose would have been similar to splitting on age as a result of the close association between age and sedoheptulose concentration in 24-h urine samples (34). The second split was done based on age as a top determinant, thereby indicating a close association between urinary maltose, sex, and age. The cut point for age was 45 y, thus suggesting that sugar excretion patterns change with menopause in women (see Figure 3). Gluconic acid was the only possible candidate metabolite for the third and last split (see Figure 3 and Table 3); however, to prevent overfitting, no further splitting was done. Boxplots of the 2 most important sugar compounds that separate men and women and the interaction between maltose excretion, sex, and age (menopausal status) are shown in Figure 3.

Association of the urinary sugar profile with diet

A correlation analysis was performed based on 24-h urinary sugar profiles with the food consumption and nutrient intake data and a heat map generated on the basis of the Spearman rank correlation coefficients (**Figure 4**). The Spearman rank correlation coefficients with $\rho > 0.30$ are listed in **Table 4**; no correlations with $\rho < -0.30$ were observed. Significant correlations were observed for *1)* avocado consumption with perseitol, *2)* dairy product consumption with galactose and lactose, *3)* alcoholic beverage consumption with xylitol and ethyl-β-glucuronide, *4)* mushroom consumption with mannitol,

Sugar compounds	Median	Minimum	Maximum	CV
Mannoheptulose	1.25E+05	3.66E+04	2.45E+07	665.2
Perseitol	1.42E+03	3.65E+02	2.32E+06	653.8
Ethyl-β-glucuronide	4.18E+05	1.79E+05	6.55E+06	328.4
Lactose	9.53E+05	4.18E+05	7.81E+06	294.5
Unknown U25	2.94E+06	1.77E+05	8.50E+05	255.0
Sucrose	4.61E+04	8.49E+03	3.36E+07	227.2
Disaccharide U28	4.55E+04	7.54E+03	2.97E+05	224.9
meso-Erythritol	9.79E+04	4.49E+04	1.66E+07	186.6
Xylitol	2.91E+04	6.06E+03	2.30E+06	181.4
Maltose	1.29E+05	1.09E+04	2.75E+05	162.4
Psicose	7.31E+04	5.44E+03	5.09E+06	162.4
Levoglucosan	4.22E+05	1.07E+05	2.48E+06	146.6
chiro-Inositol	2.32E+05	4.80E+04	7.23E+05	146.5
1-Deoxy-sorbitol	6.76E+05	1.51E+05	1.18E+06	127.2
myo-Inositol	2.34E+04	1.76E+03	5.06E+06	119.4
Mannitol	1.90E+05	4.13E+04	4.45E+07	118.7
Unknown U04	1.11E+04	9.15E+02	2.73E+05	109.0
Fructose	1.23E+04	2.46E+03	2.35E+06	100.4
Sorbitol	6.36E+04	4.30E+03	3.39E+06	92.2
Galactose	5.55E+05	1.44E+04	7.32E+05	83.3
Disaccharide U29	9.08E+04	1.34E+04	1.34E+05	80.9
Galactonic acid	1.04E+05	2.54E+04	3.80E+05	79.9
Polyol U02	3.08E+03	5.40E+02	2.81E+05	79.5
Xylose	1.40E+05	8.91E+03	4.07E+05	77.7
Allose	1.86E+05	1.29E+04	1.63E+06	73.8
Mannose	5.86E+05	2.08E+05	4.02E+05	69.9
Fucose	9.62E+04	1.00E+04	5.69E+06	60.2
Galactitol	9.09E+05	6.68E+04	1.22E+06	58.9
Sugar acid U01	3.60E+05	1.61E+05	3.09E+05	56.0
Disaccharide U26	2.46E+04	1.28E+03	2.63E+05	55.7
Sugar acid U06	3.68E+05	3.60E+04	5.20E+04	54.4
Threitol	3.94E+05	8.53E+04	5.13E+05	54.3
scyllo-Inositol	2.68E+05	3.20E+04	2.37E+06	52.6
Sugar acid U09	2.18E+04	2.22E+03	4.20E+05	51.3
Disaccharide U27	1.11E+04	1.93E+03	9.72E+05	51.0
Unknown U24	2.59E+04	3.03E+03	1.39E+06	50.9
Xylonic acid	4.84E+05	2.24E+04	3.90E+05	50.4
Unknown U12	2.41E+05	4.89E+04	1.45E+05	49.6
Arabinose	4.92E+04	1.99E+03	1.50E+06	46.0
Fucitol	1.31E+04	4.37E+03	7.82E+04	45.5
Sedoheptulose	1.70E+05	3.36E+03	2.53E+05	45.5
Gluconic acid	2.58E+05	6.13E+04	3.47E+05	42.4
Arabonic acid	8.15E+04	2.67E+04	1.03E+06	41.0
Arabitol	2.05E+04	7.54E+03	1.24E+06	38.2
Glucose	1.14E+05	8.81E+03	3.16E+06	36.4
Ribitol	8.70E+04	5.31E+03	2.85E+05	35.0
Unknown U03	7.63E+05	2.40E+05	1.38E+05	34.5
Threonic acid	5.67E+04	7.21E+02	1.14E+06	34.3
Mannonic acid	3.53E+04	4.86E+02	1.50E+05	31.9
Ribonic acid	7.41E+04	2.76E+04	1.88E+05	30.3
Monosaccharide U21	9.80E+04	1.55E+04	4.88E+04	30.0
Glucuronic acid	8.62E+03	1.18E+03	2.03E+06	29.8
Erythronic acid	2.72E+05	3.42E+04	2.03E+06	28.6
Unknown U05	4.85E+03	1.22E+03	1.41E+06	28.5
Ribose	1.20E+05	3.91E+04	2.20E+05	25.6

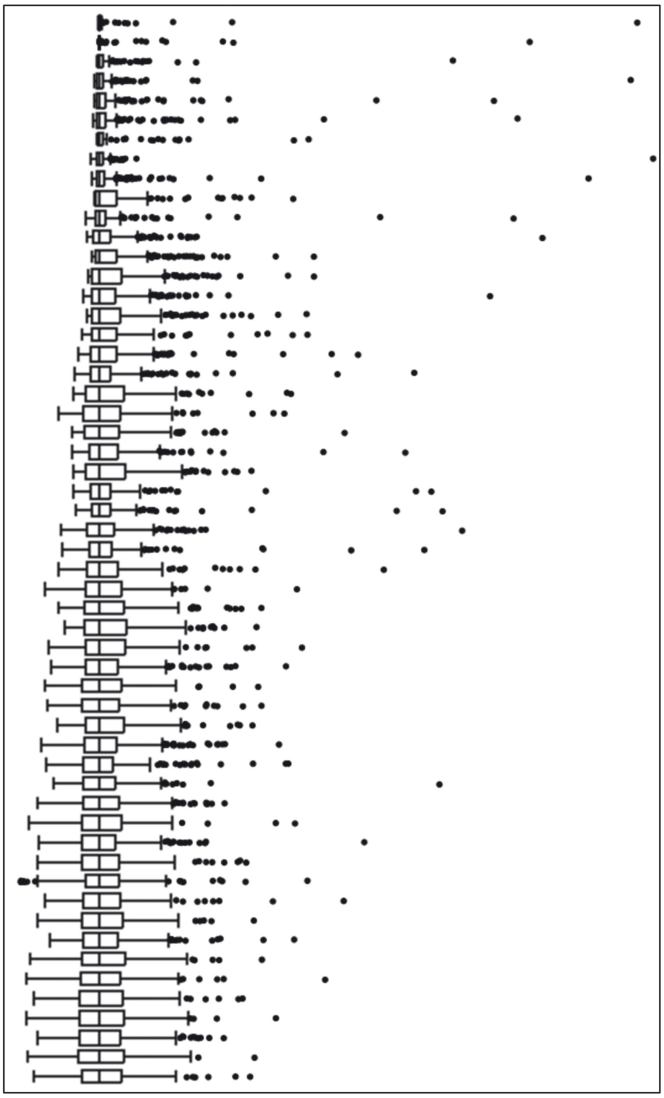


FIGURE 2 Interindividual/biological variability of the sugar profile in human urine from 301 participants of the KarMeN study. Sugar compound intensities were median-centered and unit variance scaled to allow for a comparison of the biological variance. Whiskers of box plots indicate ± 1.5 IQR. Median, minimum, maximum, and CV refer to peak area and enable a relative comparison of signal intensity. For reasons of simplicity, only the first of 2 derivatives was listed.

5) fruit consumption with threitol, xylose, and an unknown polyol, *6)* citrus fruit juice and fruit drink consumption with *chiro*-inositol and galactonic acid, and *7)* sucrose intake with fructose and sucrose (see Table 4 and Figure 4). In the case of avocado, in addition to perseitol, mannoheptulose presented itself as a potential dietary marker although the correlation coefficient was slightly below our threshold of 0.30 ($\rho = 0.2704$; see Figure 4). To verify this observation despite the low number of avocado consumers ($n = 9$), Spearman rank correlation coefficients were calculated for the avocado consumers and 18 randomly chosen nonconsumers ($n = 27$) (mannoheptulose: $\rho = 0.7748$, perseitol: $\rho = 0.8713$; see **Figure 5**). For some of these potential dietary markers box and scatter plots as well as their origin and potential confounders or other interferences are shown in Figure 5. A second line of evidence that those metabolites may be potential dietary markers for distinct foods/products was provided by analyzing the sugar profiles of 75 selected fruit and vegetable varieties (see Table 2).

DISCUSSION

Sugar profiling in participants of the KarMeN study

Sugar profile of human urine and biological variability

With the analytical method described here, we provide a straightforward and reliable tool to obtain sugar profiles and

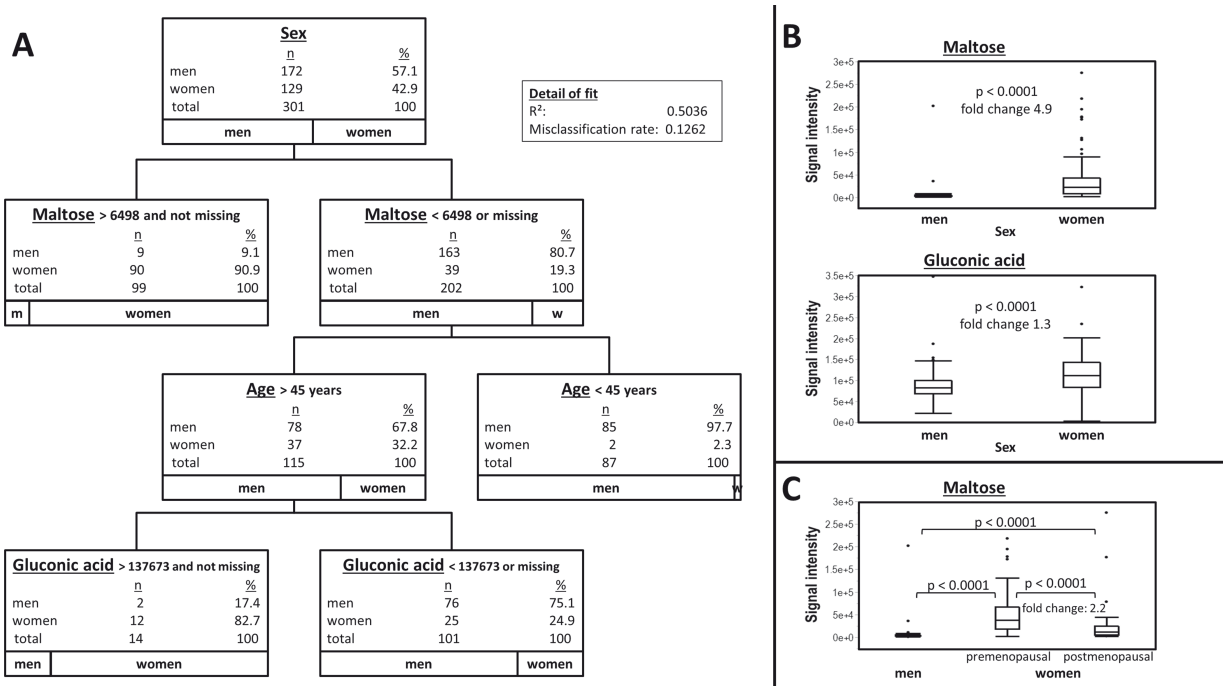


FIGURE 3 Identification of urinary markers discriminating sex via the CART approach. (A) Decision tree with splitting rules, the number of men or women, and the ratio between men and women for each branch; (B) box plots for the 2 top sugar compound candidates to differentiate sex; (C) association of age with maltose excretion in women (pre- and postmenopausal). Significance was established with the use of the Wilcoxon test, with participants excluded where the sugar compound was not detected. CART, classification and regression tree.

TABLE 3
Results of building a decision tree (CART) for the identification of possible markers to differentiate sex. Possible candidate sugar compounds for splitting are listed only if significant *P* values were achieved after Bonferroni correction[1]

Sugar compound	Candidate G^2	$-\log_{10}$ (*P* value)	Cut point[2]
First node: split candidates			
Maltose	152.6	55.8316	6.50E + 03
Gluconic acid	56.6	17.0557	1.05E + 05
Unknown U05	37.8	9.1285	6.66E + 05
Fructose[3]	27.1	5.0777	2.25E + 05
Second node: split candidates			
Age	34.7	8.6471	45.363
Gluconic acid	27.1	5.8762	1.38E + 05
Sedoheptulose[3]	25.3	5.2497	1.32E + 05
Unknown U05	19.8	3.4428	7.07E + 05
Third node: split candidates			
Gluconic acid	20.0	4.0476	1.38E + 05
First leaf: split candidates[4]			
Second leaf: split candidates[4]			
Third leaf: split candidates			
Mannonic acid	21.6	4.6179	5.62E + 04
Unknown U03	17.5	3.3893	4.73E + 04
Fourth leaf: split candidates[4]			

[1] G^2, likelihood ratio chi-square; highest values indicate best split.
[2] Best value for splitting the variables (cut point).
[3] For reasons of readability, only the higher-ranking derivative was listed.
[4] Candidate *P* values were below the significance level.

semiquantitative data in biofluids like urine. We observed a considerably wider range of sugar compounds than commonly known or expected in human urine (see Figure 2, Table 1, and Supplemental Table 1). There are many less known urinary sugar compounds with unclear origin such as psicose, perseitol, or mannoheptulose; their origin could be exogenous or endogenous. However, urinary sugar profiles partly showed a remarkable variability (see Figure 2), and thus, we believe that many individual sugars are dependent on sex, health status, or are a surrogate of a dietary pattern or the consumption of distinct food items.

Association of the urinary sugar profile with sex

The most important metabolites to differentiate between male and female sugar profiles were maltose and gluconic acid. Maltose has been reported to be present in very low concentrations in human urine, but no differences with respect to sex have been described so far (10, 12, 23, 36–39). We hypothesize that the maltose excretion seen in women may be associated with the vaginal microbiota (dominated by lactic acid–producing *Lactobacillus* species). Spear et al. (40) demonstrated that vaginal fluid possesses α-amylase activity, and thus is able to degrade free glycogen to maltose, maltotriose, and maltotetraose, which can then be utilized by *Lactobacillus* species (41, 42). This degradation pathway of free glycogen released from the vaginal epithelium might be responsible for the higher excretion rate of maltose in female urine.

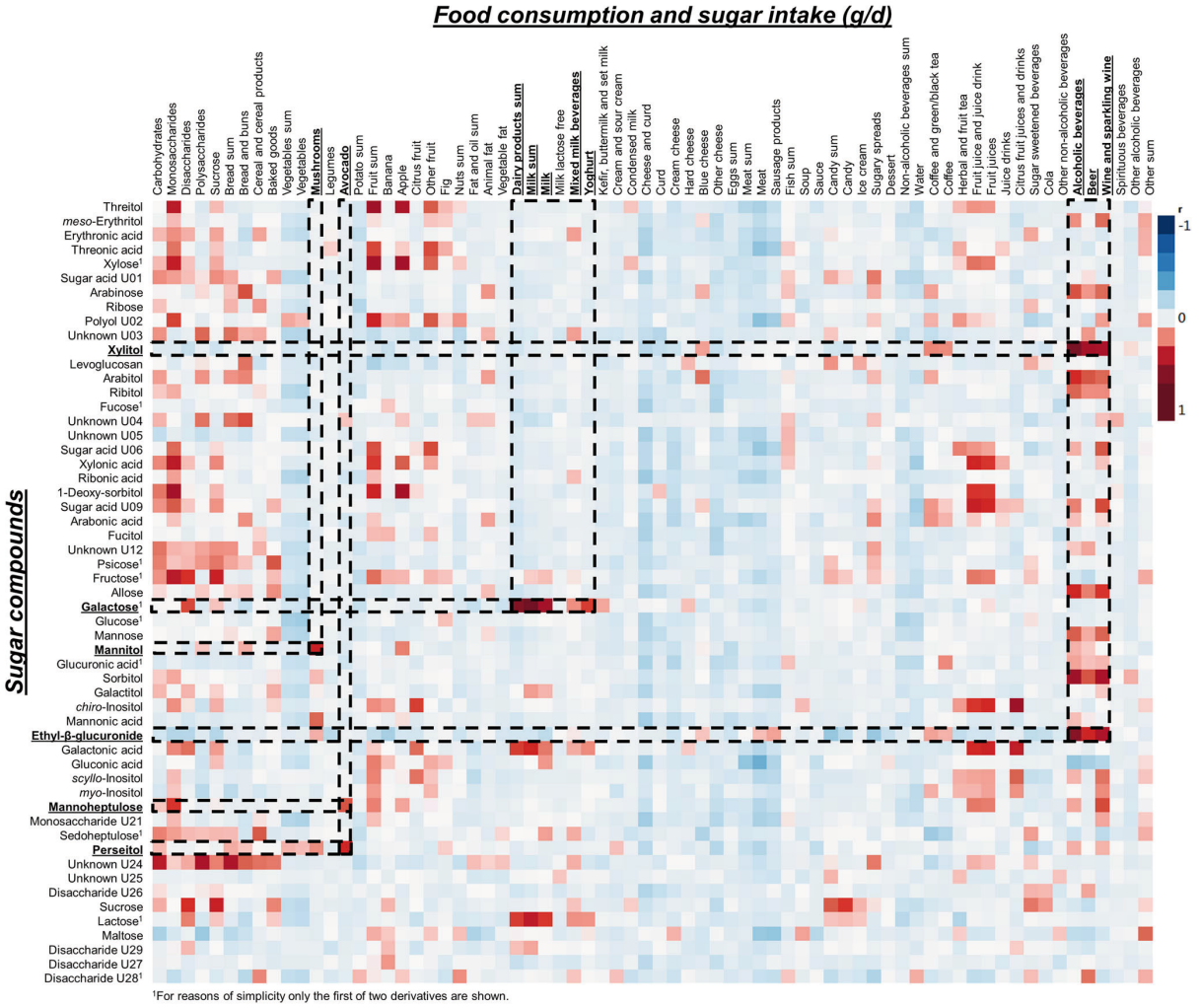


FIGURE 4 Heat map of Spearman rank correlation coefficients for the correlations of sugar compound excretion with food consumption and sugar intake (for an explanation of variables, see Supplemental Table 2) during the 24-h collection of urine. Correlations were estimated by a pairwise method. Boxes indicate the highest (most significant) correlations. The amount of consumed food is always given in g/d.

We observed a significantly lower maltose content in the urine of postmenopausal women in comparison with premenopausal women (see Figure 3). Postmenopausal women have significantly lower amounts of glycogen and *Lactobacilli* counts in the vaginal fluid as a result of reduced estrogen concentrations (41, 43, 44). Collectively, these observations concur with our finding of reduced maltose concentrations in postmenopausal women and add plausibility to a link between maltose excretion and the vaginal microbiota.

For gluconic acid and fructose we could not find a plausible biological explanation for the observed sex-dependent differences in urine.

Association of the urinary sugar profile with diet

Based on the correlation analysis, potential dietary markers for the consumption of various food items as well as food groups were identified (Figure 5 and Table 4). This analysis suggests the following sugar compounds to serve as specific dietary markers: mannoheptulose and perseitol for avocado and galactose and lactose for dairy products. These sugars are known constituents of these respective foods (45–50). Although an increase of mannoheptulose and perseitol excretion in urine after avocado consumption (25) has been described in an intervention with 3 volunteers, it was not specifically defined as a dietary marker. An increase of galactose and lactose after pure lactose ingestion has been observed (49, 51–53). Moreover, in a recent intervention study with milk, both galactose and lactose were suggested as specific dietary markers for milk consumption (54).

Although mannitol appears to be a plausible dietary marker for mushroom consumption (55), its specificity is questioned because there are many other sources of mannitol in the human diet (48) (see also Table 2) — a fact that confounds the aforementioned identified association (see Figure 5).

TABLE 4
Correlations between analytes and the consumed amounts of certain foods, food groups, or nutrients as determined by a 24-h recall (see Supplemental Table 2)[1]

Sugar compounds	Dietary intake	Participants		Spearman[2]
		$n_{excr.}$	$n_{ing.}$	ρ
Perseitol	Avocado	219	9	0.3388
Galactose[3]	Milk sum	301	234	0.6644
Galactose[3]	Dairy products sum	301	279	0.6082
Galactose[3]	Milk	301	174	0.4779
Galactose[3]	Yoghurt	301	84	0.3017
Lactose[3]	Milk sum	300	234	0.4204
Lactose[3]	Dairy products sum	300	279	0.3364
Lactose[3]	Milk	300	174	0.3180
Galactonic acid	Milk sum	296	234	0.3403
Galactonic acid	Dairy products sum	296	279	0.3005
Xylitol	Alcoholic beverages	301	96	0.6379
Xylitol	Beer	301	47	0.4667
Xylitol	Wine and sparkling wine	301	55	0.4531
Ethyl-β-glucuronide	Alcoholic beverages	144	96	0.5885
Ethyl-β-glucuronide	Wine and sparkling wine	144	55	0.4446
Ethyl-β-glucuronide	Beer	144	47	0.3457
Sorbitol	Alcoholic beverages	299	96	0.4948
Sorbitol	Wine and sparkling wine	299	55	0.4496
Allose	Alcoholic beverages	301	96	0.3425
Allose	Wine and sparkling wine	301	55	0.3128
Arabitol	Alcoholic beverages	301	96	0.3340
Mannitol	Mushrooms	301	35	0.3633
Mannitol	Button mushroom	301	32	0.3489
chiro-Inositol	Citrus fruit juices and drinks	260	65	0.4941
chiro-Inositol	Citrus fruit	260	37	0.3000
Galactonic acid	Citrus fruit juices and drinks	296	65	0.3825
Threitol	Fruit sum	301	228	0.4904
Threitol	Apple	301	106	0.4359
Xylose[3]	Fruit sum	301	228	0.4768
Xylose[3]	Apple	301	106	0.4736
1-Deoxy-sorbitol	Apple	299	106	0.4653
1-Deoxy-sorbitol	Fruit sum	299	228	0.3504
1-Deoxy-sorbitol	Fruit juice and juice drink	299	131	0.3130
1-Deoxy-sorbitol	Fruit juices	299	123	0.3097
Sugar acid U09	Fruit juice and juice drink	301	131	0.3649
Sugar acid U09	Fruit juices	301	123	0.3395
Polyol U02	Fruit sum	301	228	0.3516
Galactonic acid	Fruit juices	296	123	0.3498
Galactonic acid	Fruit juice and juice drink	296	131	0.3320
Xylonic acid	Fruit juice and juice drink	301	131	0.3408
Xylonic acid	Fruit juices	301	123	0.3107
Xylonic acid	Fruit sum	301	228	0.3091
chiro-Inositol	Fruit juices	260	123	0.3350
chiro-Inositol	Fruit juice and juice drink	260	131	0.3119
1-Deoxy-sorbitol	Monosaccharides	299	301	0.4676
Unknown U24	Polysaccharides	301	301	0.4477
Unknown U24	Bread sum	301	293	0.4463
Unknown U24	Carbohydrates	301	301	0.3949
Fructose[3]	Monosaccharides	301	301	0.3966
Fructose[3]	Sucrose	301	301	0.3581
Fructose[3]	Disaccharides	301	301	0.3192
Threitol	Monosaccharides	301	301	0.3955
Xylonic acid	Monosaccharides	301	301	0.3919
Xylose[3]	Monosaccharides	301	301	0.3785
Sucrose	Sucrose	301	301	0.3620
Sucrose	Disaccharides	301	301	0.3243
Sucrose	Candy	301	104	0.3132
Mannoheptulose	Monosaccharides	300	301	0.3135

[1] $n_{excr.}$, number of participants who excreted a certain sugar compound; $n_{ing.}$, number of participants who ingested individual foods or food groups.
[2] Spearman rank correlation coefficients < -0.30 or > 0.30. All listed correlations had significant P values < 0.0001.
[3] For reasons of readability, only the higher-ranking derivative was listed.

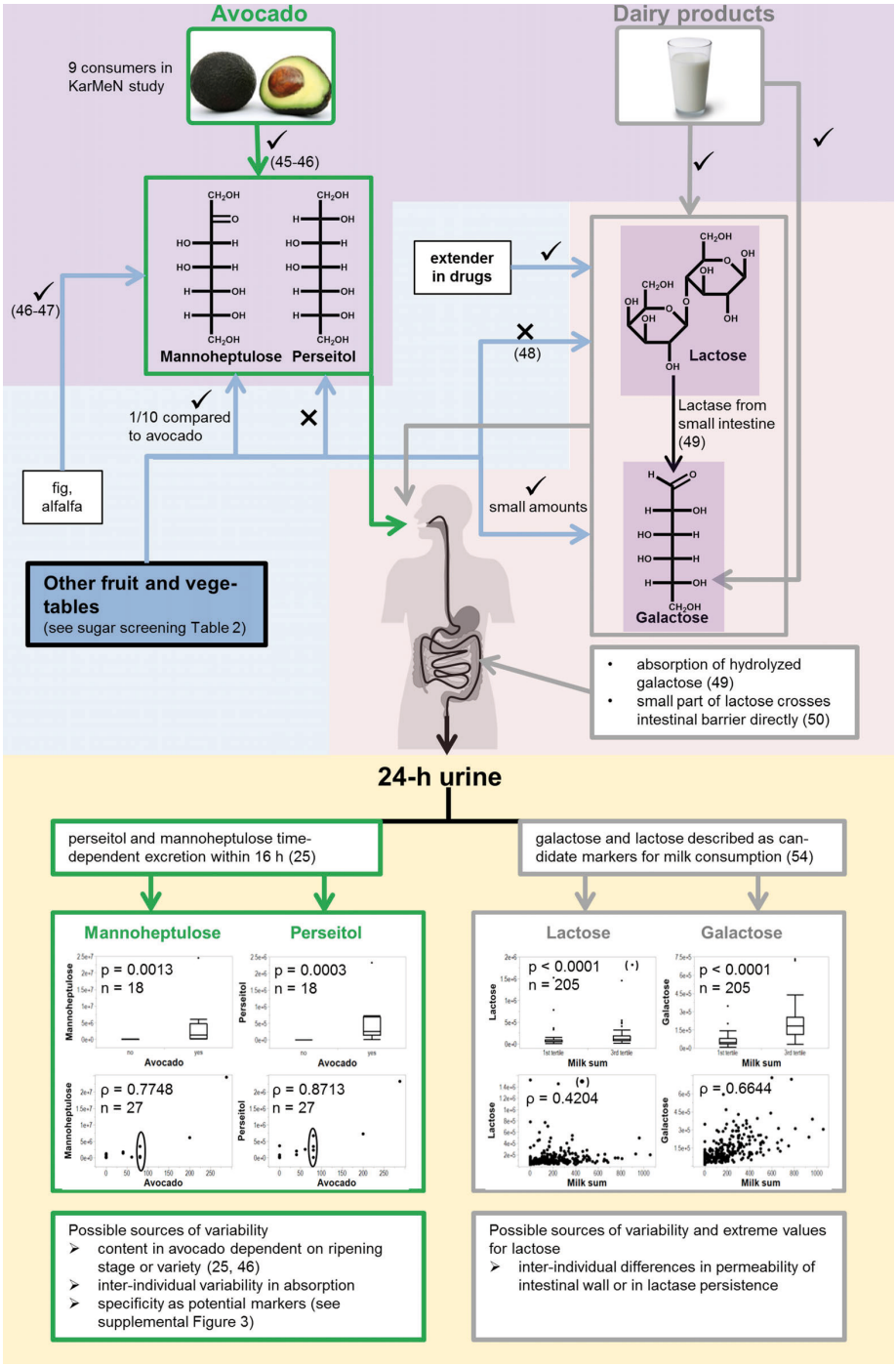


FIGURE 5 (*Continued*)

We also identified xylitol as a potential dietary marker for alcoholic beverage consumption. An increase of xylitol in urine after administration of ethanol (56) has been described before, but the causality underlying the relation between alcohol consumption and urinary xylitol output (56, 57) warrants further research. Ethyl-β-glucuronide has already been described as a dietary marker for alcoholic beverage consumption (58–60); we observed a moderate association (see Figure 5 and Table 4). In

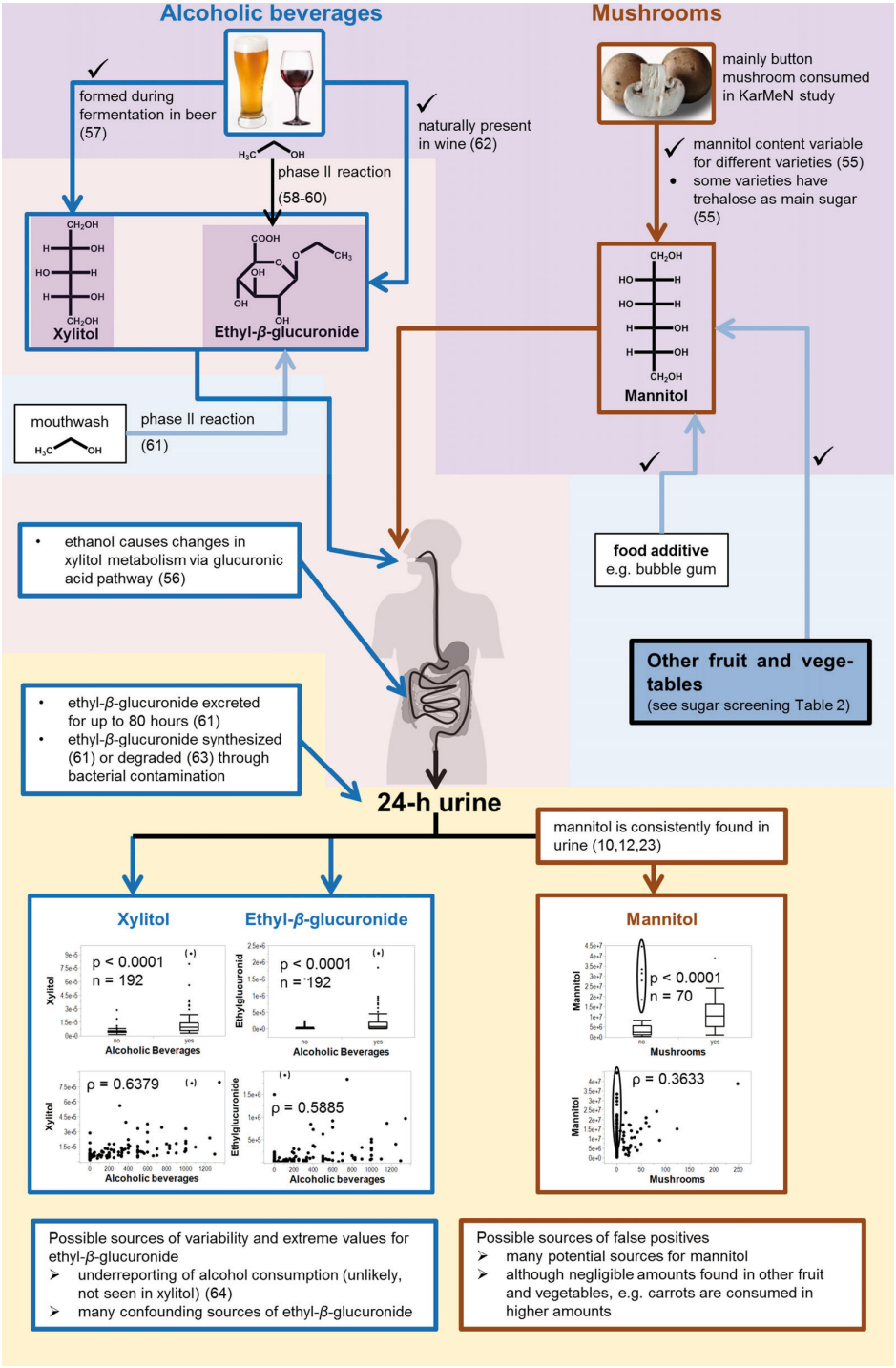


FIGURE 5 Overview of potential dietary markers (includes results of the KarMeN study and sugar screening of fruit and vegetables as well as literature data). Potential dietary markers of food consumption with the strongest associations in the correlation analysis, their plausibility in terms of origin, and their specificity in terms of potential confounders or other interferences. Shaded violet: food/nutrient level; shaded light blue: interfering sources for potential dietary markers (foods, drugs, results from sugar screening of fruit and vegetables); shaded red: metabolization in human; shaded yellow: results in 24-h urine samples. Colors of boxes and arrows as follows: green (part 1): consumption of avocado; grey (part 1): consumption of dairy products; dark blue (part 2): consumption of alcoholic beverages; brown (part 2): consumption of mushrooms; light blue: interfering sources for potential dietary markers (foods, drugs, results from sugar screening of fruit and vegetables). Check: compound occurs in specific food; cross: compound does not occur in specific food. The amount of consumed food is given in g per day. Significance was established by use of the Wilcoxon test. Spearman rank correlation coefficients were calculated using all 301 participants, except for avocado, where the 9 avocado consumers plus 18 randomly selected participants were used.

light of the many other potential confounders and interferences for ethyl-β-glucuronide detection (61–64) (see Figure 5), we recommend to use measurements of additional metabolites such as ethylsulfate (65) or in combination with xylitol.

It would be highly desirable to use some of these dietary markers in future as an objective measure of food consumption in comparison with self-reported consumption, where biases such as under- or over-reporting in cases of perceived unhealthy or healthy foods often occur (64). Objective dietary markers would allow more reliable insights into health aspects, and thus, relations between diet and health could be more accurately ascertained.

In more general terms, a dietary marker should fulfil a number of criteria such as its specificity, the dose-response relation, plausibility of origin, and suitability in free-living populations, and, importantly, analytical robustness (66). Questions around the quality of dietary markers also cover issues on whether a metabolite is a short-term marker of intake over a 24–36 h period or whether it can also serve as a long-term reporter molecule especially in epidemiologic studies (67). Moreover, whether there are saturation effects and whether the dietary marker can quantitatively assess consumption are also important issues (66).

Specificity and dose-response effect, plausibility, and suitability in a free-living population as in our KarMeN population on an unrestricted diet as well as methodological validity were all addressed in the present study. The main limitations in our approach were *1)* the low number of participants consuming some specific food items such as avocado, *2)* the potential bias through the use of self-reported food consumption data for the correlation analysis, and *3)* owing to our study design so far only a conclusion about metabolites' usefulness as short-term markers can be drawn. Other limitations might be that only a single urine collection was measured potentially leading to exaggerated interindividual variation and that only fruit and vegetables, but not processed food and beverages, were screened during the sugar profiling of food. However, our aim is that the developed analytical method and the approaches used to identify some crucial sugar compound determinants will be taken into larger and more diverse cohorts as the next step to deriving quantitative dietary markers and to shedding light on the diet-health relation for one of the most important food substrates in the human diet and metabolism, namely the sugars.

In conclusion, we have demonstrated that the human urinary sugar profile is complex and comprises many more compounds than previously anticipated. With the large number of sugar compounds detected, we identified also a huge gap in knowledge regarding the metabolism of most of these sugar compounds, in particular along the diet-health-disease trajectory. We therefore suggest that future research should not only encompass analyzing common and well-known sugar compounds, but rather strive for a more comprehensive view on sugar compounds. However, the data from our study can be used as a reference for normal sugar profiles of healthy humans with respect to the occurrence of individual sugar compounds along with variances in excretion. For some sugars, we identified crucial determinants such as sex and pre- compared with postmenopausal women. However, these determinants need further study. We also identified a considerable number of sugar compounds as potential dietary markers for individual food items and groups (see Figures 4 and 5), for which confirmation and assessment of their quantitative dimension and their usability as long-term markers in epidemiologic studies are required in future studies. Although our newly developed semitargeted GC-MS method is only semiquantitative, it clearly offers a rapid and cost-effective strategy to obtain comprehensive insights into the sugar profile by detecting not only numerous known, but also some unknown sugar-like compounds that also deserve identification. Our analytical method may also be useful in identifying the underlying physiologic processes that allow assessing determinants for absorption/permeation from the intestine into blood circulation as well as for renal secretion/reabsorption. Ultimately, this analytical method may not only help to identify dietary markers, but also to identify disease biomarkers in the future.

We thank the technicians of the GC-MS metabolomics laboratory at MRI (Silvia Remmert, Claudia Hoffmann, and Michael Meyer) and the staff of the Human Nutrition Study Center for their excellent work. We also thank Eva Hummel, Friederike Wittig, Yvonne Martiniak, and Kimberley Hoffmann for the dietary assessment of the 24-h recalls. We thank Diana Bunzel, Eva Hummel, Benedikt Merz, and Manuela Rist for proofreading the manuscript. Nicole Hillebrand is acknowledged for help with preparing the figures.

The authors' responsibilities were as follows—SEK and HD: designed the research project; AB, BW, and IH: developed the KarMeN concept and design; CM: conducted the analytical experiments; CM, BE, PGF, and CHW: analyzed data and performed statistical analysis; CM: wrote the initial draft of the manuscript; CHW, SEK, BW, IH, and HD: critically reviewed and contributed to the manuscript; CM and SEK: had primary responsibility for final content; and all authors: read and approved the final version of the manuscript. None of the authors has a conflict of interest to declare.

REFERENCES

1. Guertin KA, Moore SC, Sampson JN, Huang WY, Xiao Q, Stolzenberg-Solomon RZ, Sinha R, Cross AJ. Metabolomics in nutritional epidemiology: identifying metabolites associated with diet and quantifying their potential to uncover diet-disease relations in populations. Am J Clin Nutr 2014;100(1):208–17.
2. Playdon MC, Moore SC, Derkach A, Reedy J, Subar AF, Sampson JN, Albanes D, Gu F, Kontto J, Lassale C, et al. Identifying biomarkers of dietary patterns by using metabolomics. Am J Clin Nutr 2017;105(2):450–65.
3. Pallister T, Jennings A, Mohney RP, Yarand D, Mangino M, Cassidy A, MacGregor A, Spector TD, Menni C. Characterizing blood metabolomics profiles associated with self-reported food intakes in female twins. PLoS One 2016;11(6):e0158568.
4. Lloyd AJ, Favé G, Beckmann M, Lin W, Tailliart K, Xie L, Mathers JC, Draper J. Use of mass spectrometry fingerprinting to identify urinary metabolites after consumption of specific foods. Am J Clin Nutr 2011;94(4):981–91.
5. Luceri C, Caderni G, Lodovici M, Spagnesi MT, Monserrat C, Lancioni L, Dolara P. Urinary excretion of sucrose and fructose as a predictor of sucrose intake in dietary intervention studies. Cancer Epidemiol Biomarkers Prev 1996;5(3):167–71.
6. Tasevska N, Runswick SA, McTaggart A, Bingham SA. Urinary sucrose and fructose as biomarkers for sugar consumption. Cancer Epidemiol Biomarkers Prev 2005;14(5):1287–94.
7. Tasevska N. Urinary sugars—a biomarker of total sugars intake. Nutrients 2015;7(7):5816–33.
8. Yang Q, Zhang Z, Gregg EW, Flanders WD, Merritt R, Hu FB. Added sugar intake and cardiovascular diseases mortality among US adults. JAMA Intern Med 2014;174(4):516–24.
9. Lustig RH, Schmidt LA, Brindis CD. Public health: the toxic truth about sugar. Nature 2012;482(7383):27–9.
10. Ge S-l, Wang H, Wang Z-F, Cheng S, Wang Q-J, He P-G, Fang Y-Z. Sensitive measurement of polyols in urine by capillary zone electrophoresis coupled with amperometric detection using on-column complexation with borate. J Chromatogr B Analyt Technol Biomed Life Sci 2013;915–916:39–45.

11. Kawasaki T, Akanuma H, Yamanouchi T. Increased fructose concentrations in blood and urine in patients with diabetes. Diabetes Care 2002;25(2):353–7.
12. Pitkänen E. The serum polyol pattern and the urinary polyol excretion in diabetic and in uremic patients. Clin Chim Acta 1972;38(1): 221–30.
13. Sim H-J, Jeong J-S, Kwon H-J, Kang TH, Park HM, Lee Y-M, Kim SY, Hong S-P. HPLC with pulsed amperometric detection for sorbitol as a biomarker for diabetic neuropathy. J Chromatogr B Analyt Technol Biomed Life Sci 2009;877(14–15):1607–11.
14. Yoshii H, Uchino H, Ohmura C, Watanabe K, Tanaka Y, Kawamori R. Clinical usefulness of measuring urinary polyol excretion by gas-chromatography/mass-spectrometry in type 2 diabetes to assess polyol pathway activity. Diabetes Res Clin Pract 2001;51(2):115–23.
15. Niwa T, Yamamoto N, Maeda K, Yamada K, Ohki T, Mori M. Gas chromatographic—mass spectrometric analysis of polyols in urine and serum of uremic patients: identification of new deoxyalditols and inositol isomers. J Chromatogr B Biomed Sci Appl 1983;277: 25–39.
16. Hui M, Cheung S-W, Chin M-L, Chu K-C, Chan RC-Y, Cheng AF-B. Development and application of a rapid diagnostic method for invasive candidiasis by the detection of d-/l-arabinitol using gas chromatography/mass spectrometry. Diagn Microbiol Infect Dis 2004;49(2):117–23.
17. Yager C, Wehrli S, Segal S. Urinary galactitol and galactonate quantified by isotope-dilution gas chromatography-mass spectrometry. Clin Chim Acta 2006;366(1–2):216–24.
18. Rakotomanga S, Baillet A, Pellerin F, Baylocq-Ferrier D. Simultaneous determination of gluconolactone, galactonolactone and galactitol in urine by reversed-phase liquid chromatography: application to galactosemia. J Chromatogr B Biomed Sci Appl 1991;570(2):277–84.
19. Hiatt HH. Carbohydrate metabolism in pentosuria. Ann Intern Med 1960;53(2):372–9.
20. Moolenaar SH, Knaap MSvd, Engelke UFH, Pouwels PJW, Janssen-Zijlstra FSM, Verhoeven NM, Jakobs C, Wevers RA. In vivo and in vitro NMR spectroscopy reveal a putative novel inborn error involving polyol metabolism. NMR Biomed 2001;14(3):167–76.
21. Huck JHJ, Struys EA, Verhoeven NM, Jakobs C, van der Knaap MS. Profiling of pentose phosphate pathway intermediates in blood spots by tandem mass spectrometry: application to transaldolase deficiency. Clin Chem 2003;49(8):1375–80.
22. Huck JH, Verhoeven NM, van Hagen JM, Jakobs C, van der Knaap MS. Clinical presentations of patients with polyol abnormalities. Neuropediatrics 2004;35(3):167–73.
23. Wamelink MMC, Smith DEC, Jakobs C, Verhoeven NM. Analysis of polyols in urine by liquid chromatography–tandem mass spectrometry: a useful tool for recognition of inborn errors affecting polyol metabolism. J Inherit Metab Dis 2005;28(6):951–63.
24. Wamelink MMC, Struys EA, Jakobs C. The biochemistry, metabolism and inherited defects of the pentose phosphate pathway: a review. J Inherit Metab Dis 2008;31(6):703–17.
25. Wamelink MM, Smith DE, Jansen EE, Verhoeven NM, Struys EA, Jakobs C. Detection of transaldolase deficiency by quantification of novel seven-carbon chain carbohydrate biomarkers in urine. J Inherit Metab Dis 2007;30(5):735–42.
26. Lin S, Yang Z, Liu H, Tang L, Cai Z. Beyond glucose: metabolic shifts in responses to the effects of the oral glucose tolerance test and the high-fructose diet in rats. Mol Biosyst 2011;7(5):1537–48.
27. Zhao X, Peter A, Fritsche J, Elcnerova M, Fritsche A, Häring H-U, Schleicher ED, Xu G, Lehmann R. Changes of the plasma metabolome during an oral glucose tolerance test: is there more than glucose to look at? Am J Physiol Endocrinol Metab 2009;296(2):E384–E93.
28. Song X, Navarro SL, Diep P, Thomas WK, Razmpoosh EC, Schwarz Y, Wang C-Y, Kratz M, Neuhouser ML, Lampe JW. Comparison and validation of 2 analytical methods for measurement of urinary sucrose and fructose excretion. Nutr Res 2013;33(9):696–703.
29. Hootman KC, Trezzi J-P, Kraemer L, Burwell LS, Dong X, Guertin KA, Jaeger C, Stover PJ, Hiller K, Cassano PA. Erythritol is a pentose-phosphate pathway metabolite and associated with adiposity gain in young adults. Proc Natl Acad Sci USA 2017;114(21): E4233–E40.
30. Bub A, Kriebel A, Dörr C, Bandt S, Rist M, Roth A, Hummel E, Kulling S, Hoffmann I, Watzl B. The Karlsruhe Metabolomics and Nutrition (KarMeN) study: protocol and methods of a cross-sectional study to characterize the metabolome of healthy men and women. JMIR Research Protocols 2016;5(3):e146.
31. Slimani N, Ferrari P, Ocke M, Welch A, Boeing H, Liere M, Pala V, Amiano P, Lagiou A, Mattisson I et al. Standardization of the 24-hour diet recall calibration method used in the European Prospective Investigation into Cancer and Nutrition (EPIC): general concepts and preliminary results. Eur J Clin Nutr 2000;54(12):900–17.
32. Slimani N, Deharveng G, Charrondiere RU, van Kappel AL, Ocke MC, Welch A, Lagiou A, van Liere M, Agudo A, Pala V et al. Structure of the standardized computerized 24-h diet recall interview used as reference method in the 22 centers participating in the EPIC project. European Prospective Investigation into Cancer and Nutrition. Comput Methods Programs Biomed 1999;58(3):251–66.
33. Hartmann BM, Heuer T, Hoffmann I. The German Nutrient Database: effect of different versions on the calculated energy and nutrient intake of the German population. J Food Compos Anal 2015;42:26–9.
34. Rist MJ, Roth A, Frommherz L, Weinert CH, Krüger R, Merz B, Bunzel D, Mack C, Egert B, Bub A et al. Metabolite patterns predicting sex and age in participants of the Karlsruhe Metabolomics and Nutrition (KarMeN) study. PLoS One 2017;12(8):e0183228.
35. Dunn WB, Wilson ID, Nicholls AW, Broadhurst D. The importance of experimental design and QC samples in large-scale and MS-driven untargeted metabolomic studies of humans. Bioanalysis 2012;4(18):2249–64.
36. Shoemaker JD, Elliott WH. Automated screening of urine samples for carbohydrates, organic and amino acids after treatment with urease. J Chromatogr B Biomed Sci Appl 1991;562(1):125–38.
37. Bhatti T, Clamp JR. Identification and estimation of monosaccharides and disaccharides in urine by gas-liquid chromatography. Clin Chim Acta 1968;22(4):563–7.
38. Nakamura H, Tamura Z. Gas chromatographic analysis of mono- and disaccharides in human blood and urine after oral administration of disaccharides. Clin Chim Acta 1972;39(2):367–81.
39. Bouatra S, Aziat F, Mandal R, Guo AC, Wilson MR, Knox C, Bjorndahl TC, Krishnamurthy R, Saleem F, Liu P, et al. The human urine metabolome. PLoS One 2013;8(9):e73076.
40. Spear GT, French AL, Gilbert D, Zariffard MR, Mirmonsef P, Sullivan TH, Spear WW, Landay A, Micci S, Lee BH, et al. Human alpha-amylase present in lower-genital-tract mucosal fluid processes glycogen to support vaginal colonization by Lactobacillus. J Infect Dis 2014;210(7):1019–28.
41. Mirmonsef P, Hotton AL, Gilbert D, Burgad D, Landay A, Weber KM, Cohen M, Ravel J, Spear GT. Free glycogen in vaginal fluids is associated with Lactobacillus colonization and low vaginal pH. PLoS One 2014;9(7):e102467.
42. Nunn KL, Forney LJ. Unraveling the dynamics of the human vaginal microbiome. Yale J Biol Med 2016;89(3):331–7.
43. Mirmonsef P, Hotton AL, Gilbert D, Gioia CJ, Maric D, Hope TJ, Landay AL, Spear GT. Glycogen levels in undiluted genital fluid and their relationship to vaginal pH, estrogen, and progesterone. PLoS One 2016;11(4):e0153553.
44. Mirmonsef P, Modur S, Burgad D, Gilbert D, Golub ET, French AL, McCotter K, Landay AL, Spear GT. Exploratory comparison of vaginal glycogen and Lactobacillus levels in premenopausal and postmenopausal women. Menopause 2015;22(7):702–9.
45. La Forge FB. D-mannoketoheptose, a new sugar from the avocado. J Biol Chem 1917;28(2):511–22.
46. Liu X, Sievert J, Lu Arpaia M, Madore MA. Postulated physiological roles of the seven-carbon sugars, mannoheptulose, and perseitol in avocado. J Am Soc Hortic Sci 2002;127(1):108–14.
47. Bean RC, Barr BK, Welch HV, Porter GG. Carbohydrate metabolism of the avocado. Arch Biochem Biophys 1962;96(3): 524–9.
48. Muir JG, Rose R, Rosella O, Liels K, Barrett JS, Shepherd SJ, Gibson PR. Measurement of short-chain carbohydrates in common Australian vegetables and fruits by high-performance liquid chromatography (HPLC). J Agric Food Chem 2009;57(2):554–65.
49. Herrinton LJ, Weiss NS, Beresford SAA, Stanford JL, Wolfla DM, Feng Z, Scott CR. Lactose and galactose intake and metabolism in relation to the risk of epithelial ovarian cancer. Am J Epidemiol 1995;141(5):407–16.
50. Laker MF. Estimation of disaccharides in plasma and urine by gas-liquid chromatography. J Chromatogr B Biomed Sci Appl 1979;163(1):9–18.

51. Sharma SK, Singhal R, Malhotra BD, Sehgal N, Kumar A. Biosensor based on Langmuir–Blodgett films of poly(3-hexyl thiophene) for detection of galactose in human blood. Biotechnol Lett 2004;26(8):645–7.
52. Shinka T, Inoue Y, Peng H, Zhen-Wei X, Ose M, Kuhara T. Urine screening of five-day-old newborns: metabolic profiling of neonatal galactosuria. J Chromatogr B Biomed Sci Appl 1999;732(2):469–77.
53. Arola H. Diagnosis of hypolactasia and lactose malabsorption. Scand J Gastroenterol 1994;29(sup202):26–35.
54. Münger LH, Trimigno A, Picone G, Freiburghaus C, Pimentel G, Burton KJ, Pralong FP, Vionnet N, Capozzi F, Badertscher R, et al. Identification of urinary food intake biomarkers for milk, cheese, and soy-based drink by untargeted GC-MS and NMR in healthy humans. J Proteome Res 2017;16(9):3321–35.
55. Kalač P. A review of chemical composition and nutritional value of wild-growing and cultivated mushrooms. J Sci Food Agric 2013;93(2):209–18.
56. Pitkanen E, Sahlstrom K. Increased excretion of xylitol after administration of glucuronolactone and ethanol in man. Ann Med Exp Biol Fenn 1968;46(2):143–50.
57. Rovio S, Sirén K, Sirén H. Application of capillary electrophoresis to determine metal cations, anions, organic acids, and carbohydrates in some Pinot Noir red wines. Food Chem 2011;124(3):1194–200.
58. Weinmann W, Schaefer P, Thierauf A, Schreiber A, Wurst FM. Confirmatory analysis of ethylglucuronide in urine by liquid-chromatography/electrospray ionization/tandem mass spectrometry according to forensic guidelines. J Am Soc Mass Spectrom 2004;15(2):188–93.
59. Wurst FM, Kempter C, Metzger J, Seidl S, Alt A. Ethyl glucuronide: a marker of recent alcohol consumption with clinical and forensic implications. Alcohol 2000;20(2):111–16.
60. Wurst FM, Kempter C, Seidl S, Alt A. Ethyl glucuronide—a marker of alcohol consumption and a relapse marker with clinical and forensic implications. Alcohol Alcohol 1999;34(1):71–7.
61. Helander A, Olsson I, Dahl H. Postcollection synthesis of ethyl glucuronide by bacteria in urine may cause false identification of alcohol consumption. Clin Chem 2007;53(10):1855–7.
62. Politi L, Morini L, Groppi A, Poloni V, Pozzi F, Polettini A. Direct determination of the ethanol metabolites ethyl glucuronide and ethyl sulfate in urine by liquid chromatography/electrospray tandem mass spectrometry. Rapid Commun Mass Spectrom 2005;19(10):1321–31.
63. Helander A, Dahl H. Urinary tract infection: a risk factor for false-negative urinary ethyl glucuronide but not ethyl sulfate in the detection of recent alcohol consumption. Clin Chem 2005;51(9):1728–30.
64. Eisinger-Watzl M, Straßburg A, Ramünke J, Krems C, Heuer T, Hoffmann I. Comparison of two dietary assessment methods by food consumption: results of the German National Nutrition Survey II. Eur J Nutr 2015;54(3):343–54.
65. Dresen S, Weinmann W, Wurst FM. Forensic confirmatory analysis of ethyl sulfate—a new marker for alcohol consumption—by liquid-chromatography/electrospray ionization/tandem mass spectrometry. J Am Soc Mass Spectrom 2004;15(11):1644–8.
66. Kuhnle GGC. Nutritional biomarkers for objective dietary assessment. J Sci Food Agric 2012;92(6):1145–9.
67. Gao Q, Pratico G, Scalbert A, Vergeres G, Kolehmainen M, Manach C, Brennan L, Afman LA, Wishart DS, Andres-Lacueva C et al. A scheme for a flexible classification of dietary and health biomarkers. Genes Nutr 2017;12:34.

Online Supporting Material

The complex human urinary sugar profile: determinants revealed in the cross-sectional KarMeN study

Carina I. Mack, Christoph H. Weinert, Björn Egert, Paola G. Ferrario, Achim Bub, Ingrid Hoffmann, Bernhard Watzl, Hannelore Daniel, Sabine E. Kulling (2018):

The complex human urinary sugar profile: determinants revealed in the cross-sectional KarMeN study

The American Journal of Clinical Nutrition 108(3): 502-516

https://doi.org/10.1093/ajcn/nqy131

Published online September 11th, 2018 by the American Society for Nutrition and Oxford University Press. Reproduced with kind permission by the publisher.

Online Supporting Material

Supplemental Table 1. Sugar compounds present in human urine as reported in the current literature.

	Monosaccharides																					Disaccharides								Polyols										
	Erythrose	Threose	Deoxyribose	Arabinose	Ribose	Xylose	Xylulose	Rhamnose	Fucose	Psicose	Fructose	Sorbose	Glucose	Galactose	Mannose	Allose	1-Deoxyglucose	2-Deoxyglucose	Deoxyhexoses	Sedoheptulose	Mannoheptulose	Sucrose	Lactose	Maltose	Gentiobiose	Cellobiose	Trehalose	Palatinose	Raffinose	4-Deoxyerythritol	Deoxythreitol	Erythritol	Threitol	2-Methylerythritol	5-Deoxyxylitol	5-Deoxyarabitol	2-Deoxyribitol	Ribitol	Arabitol	Xylitol
1968: Bhatti *et al.* (1)									X		X		X	X	X							X	X	X	X	X	X	X											
1968: Jolley *et al.* (2)				X	X	X			X		X		X	X							X	X	X	X					X											
1970: Butts *et al.* (3)				X		X			X	X	X	X	X	X								X	X																	
1972: Nakamura *et al.* (4)																						X	X	X															
1972: Pitkänen *et al.* (5)																																X	X					X	X	X
1975: Pfaffenberger *et al.* (6)				X	X	X			X				X	X	X																	X	X					X	X	X
1976: Lawson *et al.* (7)																																								
1981: Mount *et al.* (8)																																								X
1983/4: Niwa *et al.* (9, 10)							X				X		X	X			X													X	X	X	X		X	X	X		X	X
1984: McIntosh (11)																																X	X					X	X	X
1986: Jansen *et al.* (12)									X		X		X	X		X						X	X									X	X					X	X	X
1987: Burke *et al.* (13)				X							X		X									X																		
1989: Haga *et al.* (14)																																X	X					X	X	X
1991: Shoemaker *et al.* (15)					X	X					X		X	X	X							X	X	X																X
2001: Yoshii *et al.* (16)											X		X																											
2004: Huck *et al.* (17)				X	X		X						X	X																		X	X					X	X	X
2005/7: Wamelink *et al.* (18, 19)																				X	X											X	X					X	X	X
2006: Lee *et al.* (20)													X																									X	X	X
2009: Koy *et al.* (21)																																X						X	X	X
2010: Fei *et al.* (22)																																X	X					X	X	X
2012: Zhang *et al.* (23)	X	X	X		X	X		X					X					X				X		X															
2013: Ge *et al.* (24)																																X						X	X	X
2013: Bouatra *et al.* (25)				X		X	X		X		X		X	X								X	X	X								X	X	X					X	X
2015: Vasquez *et al.* (26)																																	X							
2017: Bawazeer *et al.* (27)					X	X							X	X	X				X																				
Semi-targeted sugar profiling				X	X	X			X	X	X		X	X	X	X				X	X	X	X	X								X	X					X	X	X

Online Supporting Material

Supplemental Table 1 continued.

	Polyols continued																Sugar acids																Other						
	Fucitol	6-Deoxyglucitol	6-Deoxymannitol	6-Deoxyallitol	Mannitol	Sorbitol	Galactitol	Iditol	*myo*-Inositol	*neo*-Inositol	*chiro*-Inositol	*scyllo*-Inositol	*epi*-Inositol	*cis*-Inositol	Sedoheptitol	Perseitol	4-Deoxythreonic acid	3-Deoxytetronic acid	2-Deoxytetronic acid	Erythronolacton	Erythronic acid	Threonic acid	2-Deoxypentonic acid	Xylonic acid	Arabonic acid	Ribonic acid	Glucuronic acid	Galacturonic acid	Gluconic acid	Galactonic acid	Mannonic acid	Glucaric acid	N-Acetylgalactosamin	N-Acetylglucosamin	Glucosamin	Levoglucosan	Ethylglucuronid	5-Methylthioribose	**Nr. of sugars analyzed**
1968: Bhatti *et al.* (1)																																X	X					**14**
1968: Jolley *et al.* (2)																																	X					**13**
1970: Butts *et al.* (3)																																						**10**
1972: Nakamura *et al.* (4)																																						**3**
1972: Pitkänen *et al.* (5)					X	X			X																													**8**
1975: Pfaffenberger *et al.* (6)	X				X	X	X		X																													**17**
1976: Lawson *et al.* (7)																	X	X	X			X	X				X		X			X						**8**
1981: Mount *et al.* (8)					X	X	X																															**4**
1983/4: Niwa *et al.* (9, 10)	X	X	X	X	X	X			X	X	X	X		X																									**25**
1984: McIntosh (11)	X	X			X	X	X	X	X																														**12**
1986: Jansen *et al.* (12)					X	X	X		X																														**16**
1987: Burke *et al.* (13)																																							**4**
1989: Haga *et al.* (14)	X	X			X	X	X	X	X																														**12**
1991: Shoemaker *et al.* (15)						X						X															X		X	X		X		X					**17**
2001: Yoshii *et al.* (16)						X			X																									X					**4**
2004: Huck *et al.* (17)					X																																	**11**
2005/7: Wamelink *et al.* (18, 19)					X	X	X								X	X																						**12**
2006: Lee *et al.* (20)					X	X	X		X																														**8**
2009: Koy *et al.* (21)							X																																**5**
2010: Fei *et al.* (22)					X	X	X																																**8**
2012: Zhang *et al.* (23)																											X	X	X						X	X			**15**
2013: Ge *et al.* (24)					X	X	X																																**7**
2013: Bouatra *et al.* (25)					X	X			X			X					X										X	X	X			X				X	X		**26**
2015: Vasquez *et al.* (26)																				X		X																	**3**
2017: Bawazeer *et al.* (27)																																							**4**
Semi-targeted sugar profiling	X	X			X	X	X		X		X	X				X					X	X		X	X	X	X		X	X	X				X	X		**40**

Online Supporting Material

Supplemental Table 2. Description of food groups and nutrient variables derived from 24 h dietary recall and nutrient intake data. These variables were used as such for the correlation analysis.

Food/nutrient variable	Food/nutrient variable
Carbohydrates	**Eggs sum**
Monosaccharides	**Meat sum**
Disaccharides	*Meat*
Polysaccharides	*Sausage products*
Sucrose	**Fish sum**
Bread sum	**Soup**
Bread and buns (whole meal and white, brown or multigrain)	**Sauce (including warm and cold sauces; exceptions: fruit-, vegetable-, ground meat sauce)**
Cereal and cereal products	**Candy sum**
Baked goods (pastries)	*Candy*
Vegetables sum	*Ice cream*
Vegetables	*Sugary spreads (jam, jellies, honey and chocolate spreads)*
Mushrooms	*Dessert (including pudding, semolina, tiramisu)*
Legumes	**Non-alcoholic beverages sum**
Avocado	*Water*
Potato sum	*Coffee and green/black tea*
Fruit sum	Coffee
Banana	Herbal and fruit tea
Apple	*Fruit juice and juice drink*
Citrus fruit	Fruit juices
Other fruit	Juice drinks
Fig	Citrus fruit juices and drinks
Nuts sum	*Sugar sweetened beverages*
Fat and oil sum	Cola
Animal fat (including butter, lard, fish oil)	*Other non-alcoholic beverages*
Vegetable fat (including margarine)	**Alcoholic beverages**
Dairy products sum	*Beer*
Milk sum	*Wine and sparkling wine*
Milk	*Spirituous beverages*
Milk lactose free	*Other alcoholic beverages (e.g. alcopops, cocktails)*
Mixed milk beverages	**Other sum (milk substitutes, meat substitutes, cereal substitutes, sweeteners, sugar substitutes, beverage powders/-granules, herbs, spices, vegetarian spreads, protein powder, yeast, miso)**
Yoghurt	
Kefir and buttermilk and set milk	
Cream and sour cream	
Condensed milk	
Cheese and curd	
Curd	
Cream cheese	
Hard cheese	
Blue cheese	
Other cheese	

Italic and normally written variables are part of the above written summarized variables in bold or italic.

Online Supporting Material

Supplemental Table 3. Summary of all analyzed known and unknown (U) compounds detected using the semi-targeted GC-Scan-/SIM-MS profiling method, their level of identification, occurrence of partial or total chromatographic coelution, and frequency in all KarMeN participants (including excluded 11 participants). The corresponding target and reference ions are also listed. All bold set analytes were used for the statistical analysis.

Substances	Identification Level[1]	Coelution[2]	Frequency (%)	Target ion	Reference ions	Ratios[4]	Allowance[5] (%)
L-Serine	1		100.0	218.00	204.10, 117.00	185.58, 14.32	30
L-Threonine	1	$partial_{chrom.}$	100.0	218.00	117.00, 292.15	110.31, 21.75	30
***meso*-Erythritol**	1		100.0	204.10	307.15, 117.00	71.16, 143.11	30
D-Threitol	1		100.0	204.10	117.00, 307.15	126.91, 52.92	30
Erythronic acid	1	$partial_{chrom.}$	100.0	292.15	217.10, 117.00	37.58, 31.34	30
Threonic acid	1		100.0	292.15	117.00, 217.10	57.03, 44.66	30
L-Cysteine	1		100.0	220.00	218.05	70.99	30[6]
Creatinine-enol	1		100.0	329.00	314.00, 117.00	52.63, 18.71	30
Tartaric-acid	1	$partial_{chrom.}$	99.7	292.15	117.00, 277.20	4.04, 9.81	30
L-Phenylalanine	1	$partial_{mass.}$	99.0	218.00	147.05	23.14	30[6]
D-Xylose deriv. 1	1	$partial_{chrom.}$	100.0	307.15	277.20, 117.00	21.65, 46.70	50[6]
D-Xylose deriv. 2	1	$partial_{chrom.}$	100.0	160.10	277.20, 307.15	81.76, 355.50	40[6]
Sugar acid U01	3	$partial_{chrom.}$	100.0	333.10	204.10	181.23	50[6]
D-Arabinose deriv. 1	1		100.0	307.15	160.10, 277.20	22.84, 21.86	30
D-Arabinose deriv. 2	1		100.0	307.15	277.20, 160.10	21.86, 22.84	30
D-Ribose	1	$partial_{chrom.}$	100.0	307.15	277.20, 204.10	20.96, 10.53	30
Polyol U02	3		100.0	307.15	319.20, 217.10	99.91, 107.94	30
Unknown U03	4		100.0	307.15	217.10, 277.20	161.04, 21.32	30
D-Xylitol	1	$partial_{chrom.}$	100.0	307.15	319.20	85.51	30[6]
Levoglucosan	1	$partial_{chrom.}$	99.4	204.10	333.10	24.78	30[6]
D-Arabitol	1		100.0	307.15	319.20, 204.10	88.31, 52.63	30
D-Ribitol	1	$partial_{chrom.}$	100.0	319.20	307.15, 204.10	62.41, 41.07	30
L-Fucose deriv. 1	1	$partial_{chrom.}$	100.0	117.00	160.10, 277.20	13.50, 10.62	30
Unknown U04	4	$partial_{mass.}$	71.8[3]	204.10	217.10, 117.00	50.09, 9.63	30
L-Fucose deriv. 2	1		100.0	117.00	160.10, 277.20	9.95, 11.19	30
Unknown U05	4	$partial_{chrom.}$	100.0	204.10	217.10, 117.00	476.21, 13.07	30
Sugar acid U06	3	$partial_{mass.}$	100.0[3]	333.10			[6]
Xylonic acid	1		100.0	292.15	333.10, 307.15	19.21, 14.62	30

Online Supporting Material

Ribonic acid	1		100.0	292.15	333.10, 307.15	19.21, 14.62	30
1-Deoxysorbitol	2		99.4	319.20	117.00, 217.10	154.70, 37.55	30
Sugar acid U09	3		100.0	292.15	333.10	16.91	30[6]
Arabonic acid	1	$partial_{mass.}$	100.0	292.15	333.10	14.66	30[6]
D-Fucitol	1	$partial_{mass.}$	100.0[3]	319.20	117.00, 307.15	284.39, 50.19	50[6]
Unknown U11	4		69.6	245.00	292.15, 117.00	55.78, 96.48	30
Unknown U12	4	$partial_{chrom.}$	99.4	307.15	217.10	203.03	30[6]
Isocitric acid	1	$partial_{mass.}$	100.0	245.00			[6]
Hippuric acid	1	$partial_{mass.}$	100.0	206.00	105.00	117.45	30
D-Psicose deriv. 1	1	$partial_{chrom.}$	100.0	307.15	217.10	159.68	60[6]
D-Quinic acid	1		100.0	345.00	255.00, 204.10	53.26, 10.91	30
D-Psicose deriv. 2	1		100.0	307.15	204.10, 217.10	7.08, 181.54	30
D-Fructose deriv. 1	1		100.0	307.15	205.10, 217.10	29.81, 181.54	30
Allose	1	$partial_{mass.}$	100.0	319.20	205.10	63.09	30[6]
D-Fructose deriv. 2	1	$partial_{mass.}$	100.0	307.15			[6]
D-Galactose deriv. 1	1	$partial_{chrom.}$	100.0	319.20	205.10, 217.10	48.93, 20.81	30
D-Glucose deriv. 1	1		100.0	319.20	205.10, 217.10	70.92, 54.54	30
D-Mannose deriv. 2	1	$partial_{mass.}$	99.0	319.20	205.10	135.68	50[6]
D-Galactose deriv. 2	1	$partial_{chrom.}$	97.1	319.20	205.10, 217.10	64.51, 120.70	30
D-Glucose deriv. 2	1		100.0	319.20	205.10, 217.10	73.08, 23.79	30
L-Lysine	1		100.0	156.00	318.20	9.69	30[6]
D-Mannitol	1		100.0	319.20	307.15, 205.10	16.01, 60.25	30
D-Glucuronic acid deriv. 1	1	$partial_{chrom.}$	100.0	333.10	160.10	59.35	30[6]
D-Sorbitol	1	$partial_{chrom.}$	99.4	319.20	307.15, 205.10	20.98, 64.13	30
L-Tyrosine	1	$partial_{mass.}$	100.0	220.00			[6]
D-Galactitol	1	$partial_{mass.}$	99.0	319.20	307.15, 205.10	58.94, 63.87	30
D-Glucuronic acid deriv. 2	1	$partial_{mass.}$	100.0	333.10	160.10, 217.10	32.83, 5.56	30
***chiro*-Inositol**	1	$partial_{mass.}$	86.5	318.20	204.10, 307.15	3.76, 8.21	30
Unknown U16	4		99.7	202.10	319.20, 333.10	32.18, 15.53	30
D-Mannonic acid	1	$partial_{mass.}$	100.0	333.10	319.20	17.21	30[6]
Ethyl-β-D-glucuronide	1		47.4[3]	204.10	217.10, 117.00	117.01, 7.40	30
D-Galactonic acid	1		98.1	319.20	333.10, 205.10	52.62, 93.65	30
D-Gluconic acid	1		99.4	333.10	319.20, 205.10	48.75, 40.59	30
***scyllo*-Inositol**	1		100.0	318.20	204.10, 307.15	25.73, 10.18	30
***myo*-Inositol**	1	$partial_{mass.}$	100.0	318.20	307.15, 204.10	26.29, 35.63	30
Mannoheptulose	1		99.7	319.20	205.10, 217.10	58.97, 31.71	30
Monosaccharide U21	3		100.0	319.20	205.10, 217.10	50.78, 22.92	30

Online Supporting Material

Sedoheptulose deriv. 1	1		100.0	319.20	205.10, 217.10	49.13, 33.01	30
Sedoheptulose deriv. 2	1		99.7	319.20	205.10, 217.10	51.83, 33.27	30
Perseitol	1		72.1[3]	319.20	307.15, 205.10	35.09, 57.19	30
Unknown U24	4	$partial_{chrom.}$	100.0	204.10			[6]
Unknown U25	4	$partial_{chrom.}$	94.9	204.10	217.10, 307.15	348.29, 9.50	30
Disaccharide U26	3	$partial_{chrom.}$	88.1	204.10	217.10, 319.20	37.30, 23.24	30
D-Sucrose	1		100.0	361.15	319.20, 217.10	3.46, 30.72	30
D-Lactose deriv. 1	1	$partial_{chrom.}$	99.7	361.15	204.10, 319.20	112.49, 41.05	30
D-Lactose deriv. 2	1	$partial_{chrom.}$	99.7	361.15	319.20	19.61	30[6]
D-Maltose deriv. 1	1	$partial_{chrom.}$	56.1[3]	361.15	204.10, 319.20	48.63, 3.12	30
Unknown disaccharide 29	3	$partial_{chrom.}$	90.4[3]	361.15	204.10	66.65	30[6]
Unknown disaccharide 27	3	$partial_{chrom.}$	100.0	204.10	217.10, 319.20	18.39, 4.10	30
Unknown disaccharide 28 deriv. 1	3	$partial_{chrom.}$	39.1[3]	361.15	204.10, 319.20	143.93, 91.71	30
Unknown disaccharide 28 deriv. 2	3	$partial_{chrom.}$	54.2[3]	361.15	319.20, 204.10	86.44, 175.07	30

[1]1: Identified using standard substance or in-house spectral database; 2: Identified using external spectral database; 3: Assigned substance class according to mass spectra; 4: Unknown compound. [2]$partial_{mass.}$: total chromatographic coelution, separation through masses possible; $partial_{chrom.}$: partial chromatographic coelution and separation through masses possible. [3]Trace analytes. [4]Ratios of reference ions to target ion. [5]Values allowing the ratio between target ion and reference ions to deviate from the original ratio. [6]In some cases, it became necessary to use a higher allowance and/or, to use only one or no reference ion because of overlapping peaks with similar mass spectra.

Online Supporting Material

Supplemental Table 4. Intra-day (n = 8) and inter-day (n = 140) repeatability and intermediate precision of internal standards and sugar compounds in QC samples before signal intensity drift/batch correction.

Analytes	Intra-day															Inter-day
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	mean	
Xylose	3.9	9.6	4.2	11.3	5.7	5.3	5.1	2.8	6.4	3.5	2.6	5.4	3.1	3.0	5.1	11.6
Arabinose	3.1	9.2	2.2	10.5	6.1	6.1	5.4	2.6	8.8	4.3	3.1	6.2	3.2	2.4	5.2	12.8
Ribose	2.7	9.1	2.1	8.9	5.4	5.7	5.1	2.5	8.9	2.9	2.7	6.2	2.9	2.5	4.8	12.6
Fucose	2.4	9.4	4.2	5.9	5.5	5.5	5.3	2.7	8.1	3.8	2.8	4.5	3.4	2.9	4.7	8.6
Psicose	3.1	9.2	3.4	7.7	6.0	6.1	5.9	3.3	7.7	2.6	2.9	3.4	3.3	2.7	4.8	13.7
Fructose	2.9	9.4	2.4	8.3	6.8	5.9	6.5	3.5	6.7	2.0	3.7	3.5	2.8	3.1	4.8	13.7
Allose	4.4	10.5	5.8	8.0	5.1	6.8	5.0	3.7	7.4	4.0	2.8	4.4	2.7	1.7	5.2	15.7
Galactose	2.6	10.7	5.6	9.1	5.7	6.6	5.1	3.8	7.8	4.2	2.8	3.8	2.9	2.0	5.2	16.3
Glucose	1.8	9.9	7.1	6.9	6.3	6.5	5.6	3.3	8.0	3.7	2.7	4.1	2.8	2.7	5.1	15.8
Mannose	2.0	9.7	5.7	7.2	3.9	6.3	5.2	4.8	7.2	3.4	2.3	6.6	3.0	2.9	5.0	13.8
Mannoheptulose	3.2	11.4	6.3	6.6	5.6	5.8	5.5	5.5	8.2	4.1	4.9	8.5	5.8	4.9	6.2	16.5
Sedoheptulose	3.2	10.4	5.8	6.9	7.1	5.6	4.8	2.3	6.9	2.9	2.2	3.7	3.5	2.4	4.8	15.8
Unknown monosaccharide U03	3.5	7.9	3.1	10.0	7.2	5.1	6.9	3.4	8.7	5.1	5.4	5.3	3.0	2.7	5.5	13.0
Unknown monosaccharide U21	4.3	10.9	6.8	6.1	5.5	5.2	4.5	2.7	10.0	2.0	2.4	2.8	2.9	2.4	4.9	20.1
Sucrose	1.7	12.3	2.8	5.3	9.2	8.5	5.4	6.7	13.4	4.9	6.4	10.7	4.8	6.2	7.0	15.7
Lactose	2.6	14.3	6.1	5.9	7.7	13.6	8.0	10.9	12.5	9.6	7.6	42.0	3.9	4.7	10.7	29.3
Maltose †	1.7	9.8	5.5	26.1	5.4	10.5	23.5	39.4	16.1	9.5	10.6	43.2	23.0	7.5	16.5	29.1
Unknown disaccharide U26	3.0	10.7	3.1	5.7	5.2	8.2	4.7	5.4	9.5	4.9	2.6	19.7	2.8	2.4	6.3	19.3
Unknown disaccharide U27	23.2	12.5	5.2	4.9	59.0	13.1	5.2	8.3	16.0	9.7	7.3	10.6	15.0	7.0	14.1	33.8
Unknown disaccharide U28†	5.1	11.8	4.6	3.8	9.7	6.8	4.0	4.9	14.1	3.7	2.4	13.2	5.5	2.5	6.6	16.5
Unknown disaccharide U29†	7.2	25.4	6.5	11.1	18.0	16.0	9.7	43.4	34.4	13.5	22.2	57.7	8.3	13.8	20.5	32.1
Threitol	1.7	9.0	3.4	5.1	5.1	4.9	5.3	2.2	9.2	4.5	2.6	5.9	3.7	2.4	4.6	13.4
meso-Erythritol	2.5	9.1	2.7	4.9	4.4	4.5	5.5	2.6	9.1	4.3	3.4	6.1	3.4	2.4	4.6	12.9
Xylitol	2.0	9.5	5.2	6.9	6.0	5.8	6.1	2.6	10.5	3.9	2.7	5.9	3.2	2.6	5.2	14.1
Arabitol	2.5	9.2	4.3	6.2	6.7	6.1	5.8	2.7	10.3	4.3	2.8	5.7	3.1	2.4	5.1	13.1
Ribitol	2.2	9.9	5.9	6.3	6.8	6.2	8.3	3.1	8.9	4.1	3.1	6.1	3.1	2.5	5.5	14.7
1-Deoxy-sorbitol	2.0	9.3	9.4	6.1	6.1	7.1	6.1	3.8	9.3	4.3	2.5	6.1	3.0	2.4	5.5	15.9
Fucitol †	4.5	10.3	9.9	7.0	6.8	6.4	6.7	3.5	9.1	4.5	2.0	4.7	2.7	2.7	5.8	15.9
Mannitol	1.6	9.9	7.5	6.2	7.0	6.7	5.4	4.2	10.5	4.0	2.4	4.5	3.0	2.5	5.4	14.9

Online Supporting Material

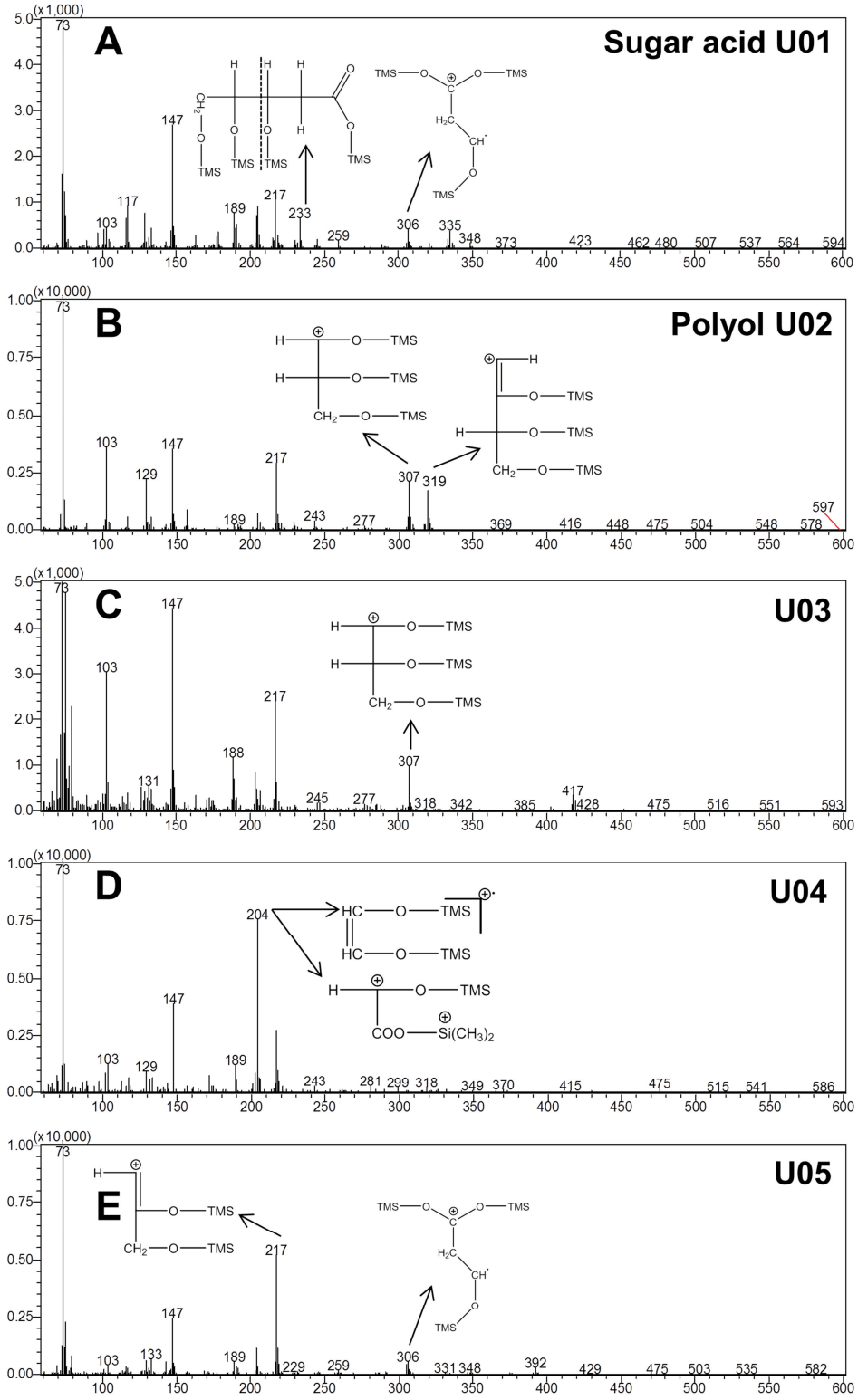
Sorbitol	2.3	8.7	10.0	7.0	7.6	6.3	6.8	4.2	10.6	4.5	2.4	7.3	4.1	4.0	6.1	13.7
Galactitol	5.2	12.4	4.7	8.7	11.3	10.8	8.1	9.6	11.8	5.0	4.2	8.1	6.1	9.3	8.2	15.7
Perseitol †	5.8	16.4	24.1	13.1	14.6	11.3	11.6	13.2	19.4	12.9	16.1	7.0	25.4	30.1	15.8	29.2
chiro-Inositol †	3.8	9.4	3.0	7.8	8.0	8.6	4.9	4.1	10.9	4.4	5.2	8.2	3.4	3.2	6.1	14.9
scyllo-Inositol	2.4	10.1	3.9	5.9	10.1	6.0	4.4	4.6	14.1	3.7	2.7	3.7	3.1	2.6	5.5	14.7
myo-Inositol	2.5	10.1	3.6	5.7	10.2	5.9	4.6	4.5	14.0	4.1	2.7	3.0	3.4	2.6	5.5	13.1
Unknown polyol U02	5.7	10.3	5.8	8.9	6.5	7.2	5.5	3.3	10.4	4.4	2.3	5.7	3.7	4.1	6.0	13.3
Erythronic acid	2.1	9.3	2.4	5.6	6.3	5.6	5.6	2.3	10.1	4.9	3.3	5.5	3.7	2.6	5.0	12.8
Threonic acid	2.0	9.0	2.8	5.7	6.2	5.6	5.0	2.4	10.3	4.1	2.9	6.1	3.5	2.6	4.9	12.4
Xylonic acid	2.7	8.7	5.0	5.2	7.2	7.8	5.2	3.7	10.1	4.7	3.4	4.4	3.0	2.6	5.3	13.5
Ribonic acid	2.2	9.9	6.0	6.3	8.2	5.0	6.3	2.8	9.8	6.5	5.6	5.6	3.1	3.5	5.8	12.9
Arabonic acid	2.4	9.8	2.4	5.3	6.0	6.1	5.2	2.6	8.6	6.0	4.3	4.1	3.7	2.6	4.9	12.3
Glucuronic acid	1.7	9.8	7.0	10.0	8.2	6.9	5.8	3.4	8.3	3.4	3.1	4.8	2.8	3.9	5.7	12.4
Mannonic acid	7.9	9.0	4.9	9.1	7.6	5.6	5.5	5.7	6.1	7.5	6.9	5.2	6.5	4.2	6.5	9.2
Galactonic acid	8.2	11.9	7.6	5.4	6.7	7.9	6.5	5.2	9.0	8.6	8.1	10.9	5.3	5.0	7.6	20.1
Gluconic acid	2.9	9.8	3.7	7.8	6.4	10.5	8.5	8.5	10.2	13.3	8.7	10.2	6.1	8.0	8.2	14.3
Unknown sugar acid U01	4.9	10.8	3.0	7.2	7.9	6.8	7.8	5.1	8.9	9.1	7.0	7.8	5.4	3.3	6.8	11.8
Unknown sugar acid U06†	5.7	12.2	12.8	10.7	9.9	7.5	8.4	4.6	10.2	7.9	6.5	13.5	7.2	6.7	8.8	14.5
Unknown sugar acid U09	2.4	9.6	3.6	5.7	7.3	6.0	5.7	2.4	8.8	4.8	3.4	3.6	2.9	2.0	4.9	13.6
Levoglucosan	2.8	11.4	4.5	4.7	5.9	4.6	4.7	4.6	7.7	7.1	4.1	5.1	4.9	5.1	5.5	11.7
Ethyl-β-D-glucuronide	2.6	9.8	2.9	3.6	3.5	8.6	3.7	4.6	6.7	5.9	3.7	8.6	3.2	3.7	5.1	16.1
Unknown U05	11.7	8.8	6.2	8.4	24.1	9.5	5.9	4.7	23.5	8.2	5.7	26.5	9.5	8.5	11.5	20.9
Unknown U12	6.6	11.2	4.5	7.0	5.6	5.5	4.9	3.0	8.9	5.0	3.7	5.3	3.6	2.6	5.5	14.4
Unknown U24	2.2	10.4	3.1	4.2	6.3	6.4	4.2	3.1	11.3	4.1	2.2	6.6	4.1	2.9	5.1	15.8
Unknown U25†	9.5	15.0	5.8	4.8	5.3	14.8	6.3	20.4	7.8	6.0	8.9	21.5	3.8	4.9	9.6	26.3
1-*O*-Methyl-2-deoxyribose	2.6	3.3	2.3	4.5	4.0	4.8	2.9	3.9	5.8	7.7	3.5	4.4	3.9	2.7	4.0	12.9
D-Pinitol	3.7	2.0	2.5	5.6	4.8	6.8	2.7	3.5	6.7	4.9	4.1	4.7	1.3	2.5	4.0	10.4
Phenyl-β-D-glucopyranosid	2.3	3.8	3.1	3.7	5.3	7.4	2.7	2.9	5.9	4.1	3.3	2.7	1.6	2.5	3.7	12.4
Mean	3.9	10.2	5.2	7.2	8.2	7.2	6.1	5.9	10.3	5.4	4.6	9.1	4.8	4.2	6.6	15.9
Median	2.7	9.8	4.6	6.3	6.4	6.4	5.5	3.7	9.1	4.5	3.3	5.9	3.4	2.7	5.5	14.3

† Trace analytes

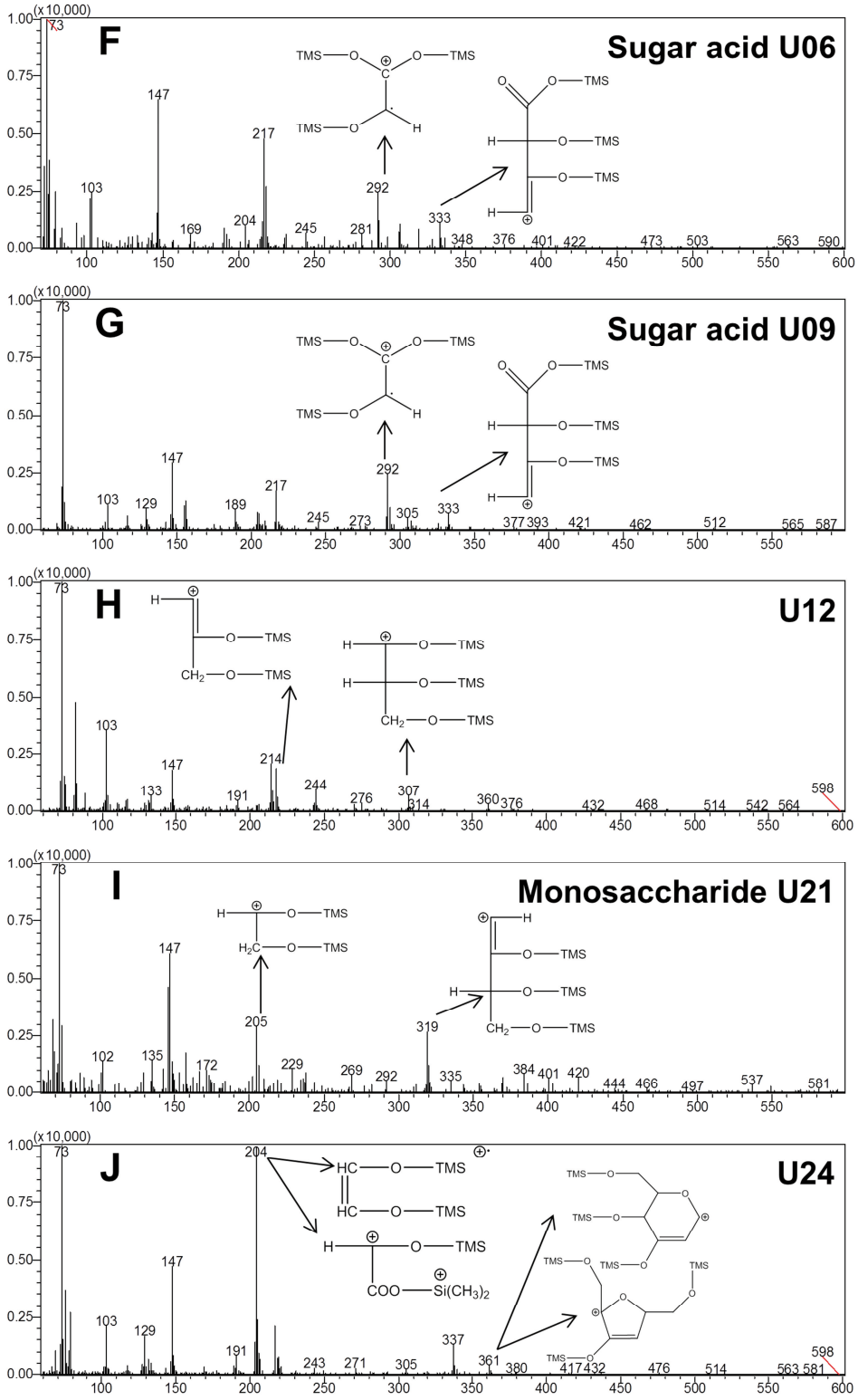
Online Supporting Material

In total, 55 known and unknown sugar compounds were detected in 24 h urine samples including analytes that were not consistently detected in all participants (see Section 3.2/3.2.1). Identification levels, frequency in study samples, and partial coelution are shown in Supplemental Table 3. To evaluate the long-term repeatability and intermediate precision during a large-scale analysis with overall 312 study samples (including the excluded 11 participants) and 140 QC samples, coefficients of variation (CV) of (uncorrected) peak areas of all sugar compounds in the QC samples were calculated (see Supplemental Table 4). In Supplemental Figure 2, the distribution of the average intra-day and inter-day repeatability is depicted. The internal standards in study and QC samples had average intra- and inter-day repeatabilities of between 2.6 and 8.2% and 10.5 and 13.4%, respectively. Despite this fact, a QC sample based drift correction was performed to remove the shallow, but steady, long-term trends in signal intensity leading to the slightly worsened inter-day repeatability (see Figure 2 and section 2.2).

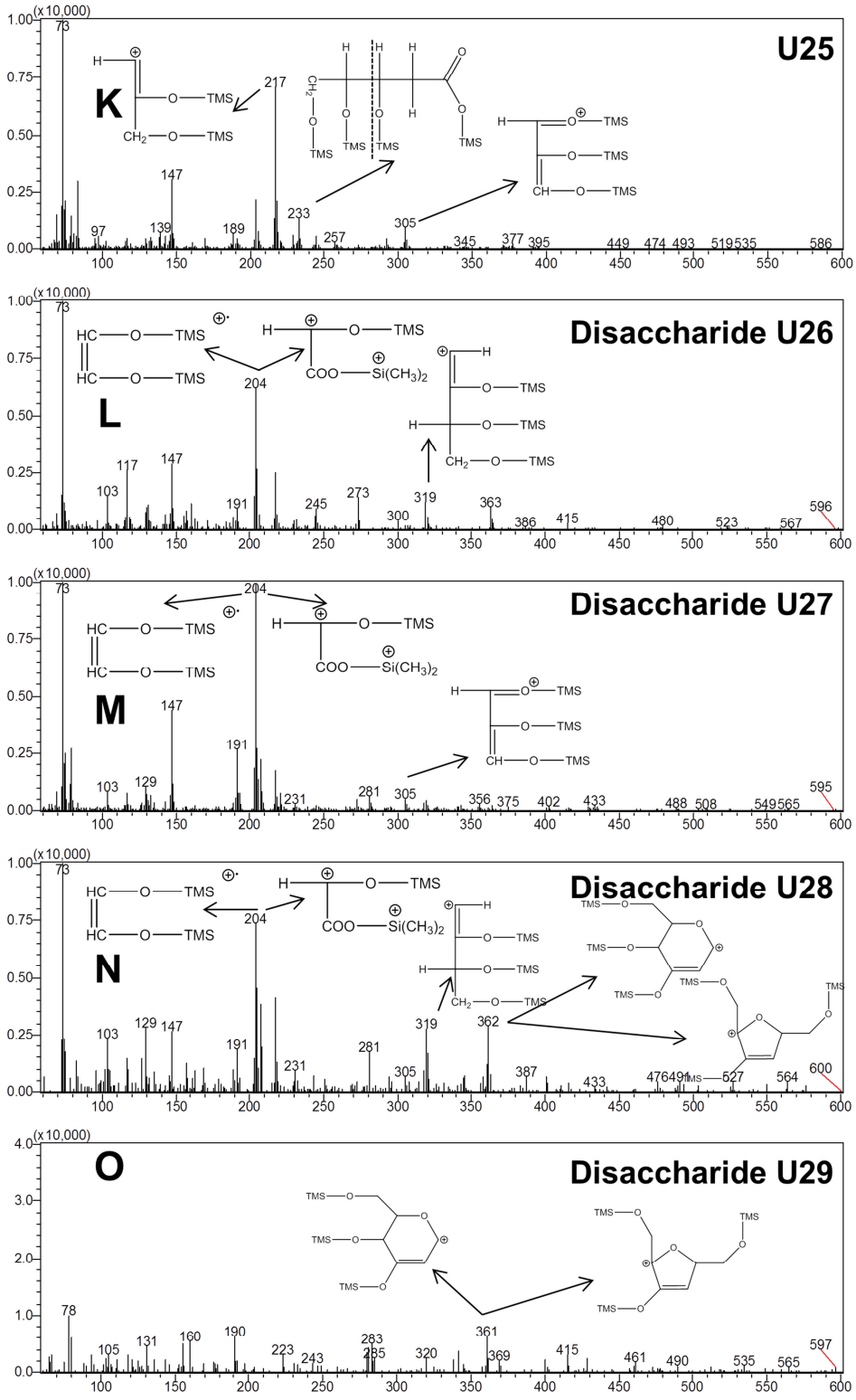
Online Supporting Material



Online Supporting Material

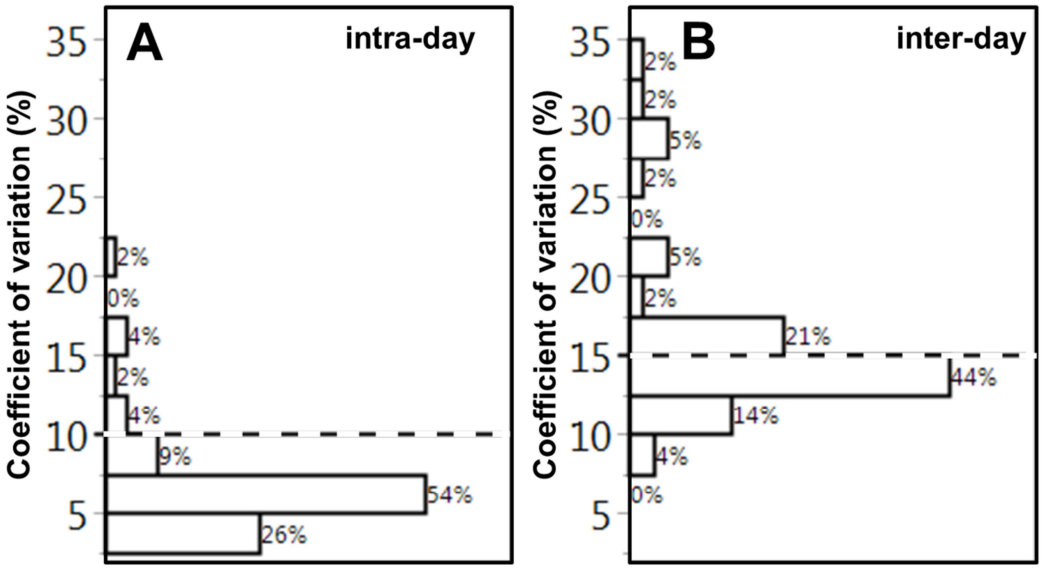


Online Supporting Material



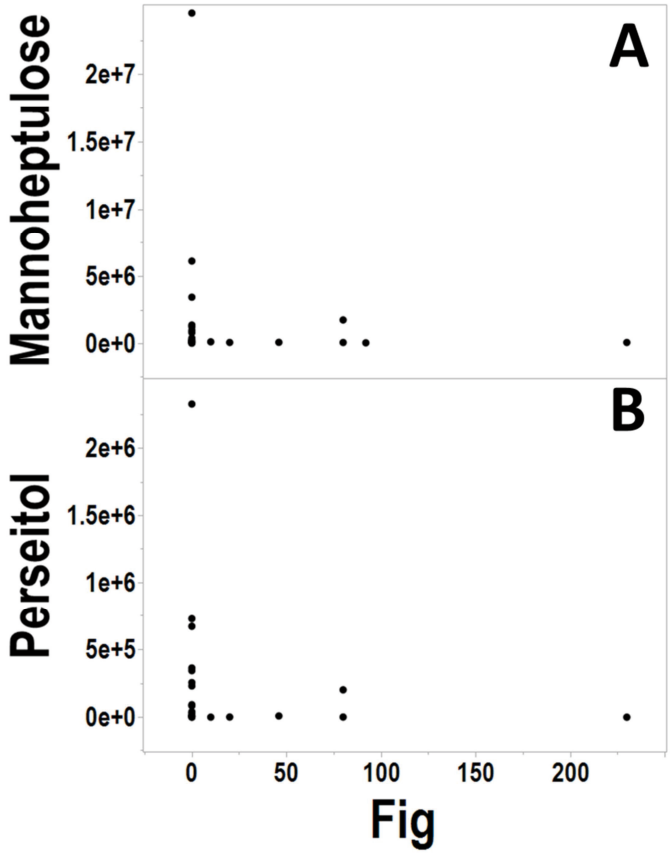
Supplemental Figure 1. Spectra of unknown sugar-like compounds (A-O) and their assignments to substance classes.

Online Supporting Material



Supplemental Figure 2. Distributions of the calculated average intra-day (A) and inter-day (B) CVs for all sugar metabolites measured in the QC samples (before drift correction). Broken lines indicate level for excellent repeatability (10%) and good repeatability (15%) for intra- (A) and inter-day (B), respectively.

Online Supporting Material



Supplemental Figure 3. Scatter plots of fig consumption (g/d) with mannoheptulose (A) and perseitol (B) excretion in urine of all 301 participants, including 7 fig consumers.

Online Supporting Material

References

1. Bhatti T, Clamp JR. Identification and estimation of monosaccharides and disaccharides in urine by gas-liquid chromatography. Clinica Chimica Acta 1968;22(4):563-7. doi: http://dx.doi.org/10.1016/0009-8981(68)90104-6.
2. Jolley RL, Freeman ML. Automated Carbohydrate Analysis of Physiologic Fluids. Clinical Chemistry 1968;14(6):538-47.
3. Butts WC, Jolley RL. Gas-Chromatographic Identification of Urinary Carbohydrates Isolated by Anion-Exchange Chromatography. Clinical Chemistry 1970;16(8):722-5.
4. Nakamura H, Tamura Z. Gas Chromatographic Analysis of Disaccharides. II. Analysis of Disaccharides in Human Body Fluids. CHEMICAL & PHARMACEUTICAL BULLETIN 1972;20(5):1070-8. doi: 10.1248/cpb.20.1070.
5. Pitkänen E. The serum polyol pattern and the urinary polyol excretion in diabetic and in uremic patients. Clinica Chimica Acta 1972;38(1):221-30. doi: http://dx.doi.org/10.1016/0009-8981(72)90230-6.
6. Pfaffenberger CD, Szafranek J, Horning MG, Horning EC. Gas chromatographic determination of polyols and aldoses in human urine as polyacetates and aldononitrile polyacetates. Analytical Biochemistry 1975;63(2):501-12. doi: http://dx.doi.org/10.1016/0003-2697(75)90374-7.
7. Lawson AM, Chalmers RA, Watts RW. Urinary organic acids in man. I. Normal patterns. Clinical Chemistry 1976;22(8):1283-7.
8. Mount JN, Laker MF. Estimation of sugar alcohols by gas—liquid chromatography using a modified acetylation procedure. Journal of Chromatography B: Biomedical Sciences and Applications 1981;226(1):191-7. doi: http://dx.doi.org/10.1016/S0378-4347(00)84220-6.
9. Niwa T, Yamada K, Ohki T, Saito A, Mori M. International symposium on high-performance liquid chromatography in the biological sciencesIdentification of 6-deoxyallitol and 6-deoxygulitol in human urine. Journal of Chromatography B: Biomedical Sciences and Applications 1984;336(1):345-50. doi: http://dx.doi.org/10.1016/S0378-4347(00)85158-0.
10. Niwa T, Yamamoto N, Maeda K, Yamada K, Ohki T, Mori M. Gas chromatographic—mass spectrometric analysis of polyols in urine and serum of uremic patients : Identification of new deoxyalditols and inositol isomers. Journal of Chromatography B: Biomedical Sciences and Applications 1983;277(0):25-39. doi: http://dx.doi.org/10.1016/S0378-4347(00)84820-3.
11. McIntosh JC. The excretion of sugars in human urine. Clinica Chimica Acta 1984;143(2):169-72. doi: http://dx.doi.org/10.1016/0009-8981(84)90225-0.
12. Jansen G, Muskiet FAJ, Schierbeek H, Berger R, van der Slik W. Capillary gas chromatographic profiling of urinary, plasma and erythrocyte sugars and polyols as their trimethylsilyl derivatives, preceded by a simple and rapid prepurification method. Clinica Chimica Acta 1986;157(3):277-93. doi: http://dx.doi.org/10.1016/0009-8981(86)90303-7.
13. Burke K, Berry C, Cooke M. Improved Method for the Determination of Sugars Using Capillary Gas Chromatography with Flame-ionisation and Nitrogen-selective Detection. Analyst 1987;112:1427-31.
14. Haga H, Nakajima T. Determination of polyol profiles in human urine by capillary gas chromatography. Biomedical Chromatography 1989;3(2):68-71. doi: 10.1002/bmc.1130030206.

Online Supporting Material

15. Shoemaker JD, Elliott WH. Automated screening of urine samples for carbohydrates, organic and amino acids after treatment with urease. Journal of Chromatography B: Biomedical Sciences and Applications 1991;562(1):125-38. doi: http://dx.doi.org/10.1016/0378-4347(91)80571-S.
16. Yoshii H, Uchino H, Ohmura C, Watanabe K, Tanaka Y, Kawamori R. Clinical usefulness of measuring urinary polyol excretion by gas-chromatography/mass-spectrometry in type 2 diabetes to assess polyol pathway activity. Diabetes Research and Clinical Practice 2001;51(2):115-23. doi: http://dx.doi.org/10.1016/S0168-8227(00)00221-7.
17. Huck JH, Verhoeven NM, van Hagen JM, Jakobs C, van der Knaap MS. Clinical presentations of patients with polyol abnormalities. Neuropediatrics 2004;35(3):167-73. doi: 10.1055/s-2004-820918.
18. Wamelink MM, Smith DE, Jansen EE, Verhoeven NM, Struys EA, Jakobs C. Detection of transaldolase deficiency by quantification of novel seven-carbon chain carbohydrate biomarkers in urine. J Inherit Metab Dis 2007;30(5):735-42. doi: 10.1007/s10545-007-0590-2.
19. Wamelink MMC, Smith DEC, Jakobs C, Verhoeven NM. Analysis of polyols in urine by liquid chromatography–tandem mass spectrometry: A useful tool for recognition of inborn errors affecting polyol metabolism. J Inherit Metab Dis 2005;28(6):951-63. doi: 10.1007/s10545-005-0233-4.
20. Lee J, Chung BC. Simultaneous measurement of urinary polyols using gas chromatography/mass spectrometry. Journal of Chromatography B 2006;831(1–2):126-31. doi: http://dx.doi.org/10.1016/j.jchromb.2005.11.043.
21. Koy A, Waldhaus A, Hammen H-W, Wendel U, Mayatepek E, Schadewaldt P. Urinary excretion of pentose phosphate pathway-associated polyols in early postnatal life. Neonatology 2009;95(3):256-61. doi: 10.1159/000167789.
22. Fei F, Britz-McKibbin P. Direct analysis of polyols using 3-nitrophenylboronic acid in capillary electrophoresis: thermodynamic and electrokinetic principles of molecular recognition. Anal Bioanal Chem 2010;398(3):1349-56. doi: 10.1007/s00216-010-4038-4.
23. Zhang T, Creek DJ, Barrett MP, Blackburn G, Watson DG. Evaluation of Coupling Reversed Phase, Aqueous Normal Phase, and Hydrophilic Interaction Liquid Chromatography with Orbitrap Mass Spectrometry for Metabolomic Studies of Human Urine. Analytical Chemistry 2012;84(4):1994-2001. doi: 10.1021/ac2030738.
24. Ge S-l, Wang H, Wang Z-F, Cheng S, Wang Q-J, He P-G, Fang Y-Z. Sensitive measurement of polyols in urine by capillary zone electrophoresis coupled with amperometric detection using on-column complexation with borate. Journal of Chromatography B 2013;915–916:39-45. doi: http://dx.doi.org/10.1016/j.jchromb.2012.12.017.
25. Bouatra S, Aziat F, Mandal R, Guo AC, Wilson MR, Knox C, Bjorndahl TC, Krishnamurthy R, Saleem F, Liu P, et al. The Human Urine Metabolome. PLoS ONE 2013;8(9):e73076. doi: 10.1371/journal.pone.0073076.
26. Vasquez NP, Crosnier de bellaistre-Bonose M, Lévêque N, Thioulouse E, Doummar D, Billette de Villemeur T, Rodriguez D, Couderc R, Robin S, Courderot-Masuyer C, et al. Advances in the metabolic profiling of acidic compounds in children's urines achieved by comprehensive two-dimensional gas chromatography. Journal of Chromatography B 2015;1002:130-8. doi: http://dx.doi.org/10.1016/j.jchromb.2015.08.006.
27. Bawazeer S, Muhsen Ali A, Alhawiti A, Khalaf A, Gibson C, Tusiimire J, Watson DG. A method for the analysis of sugars in biological systems using reductive amination in

Online Supporting Material

combination with hydrophilic interaction chromatography and high resolution mass spectrometry. Talanta 2017;166:75-80. doi: 10.1016/j.talanta.2017.01.038.

3.3 Exploring the diversity of sugar compounds in healthy, prediabetic and diabetic volunteers

Carina I. Mack, Paola G. Ferrario, Christoph H. Weinert, Björn Egert, Anja S. Hoefle, Yu-Mi Lee, Thomas Skurk, Sabine E. Kulling, Hannelore Daniel (2020):

Exploring the diversity of sugar compounds in healthy, prediabetic and diabetic volunteers

Molecular Nutrition and Food Research 2020, 64, 1901190.

https://doi.org/10.1002/mnfr.201901190

RESEARCH ARTICLE

Molecular Nutrition & Food Research
www.mnf-journal.com

Exploring the Diversity of Sugar Compounds in Healthy, Prediabetic, and Diabetic Volunteers

Carina I. Mack, Paola G. Ferrario, Christoph H. Weinert, Björn Egert, Anja S. Hoefle, Yu-Mi Lee, Thomas Skurk, Sabine E. Kulling, and Hannelore Daniel*

Scope: Diabetes is thought to primarily represent a disturbance of carbohydrate metabolism; however, population studies employing metabolomics have mainly identified plasma amino acids and lipids, or their products, as biomarkers. In this pilot study, the aim is to analyze a wide spectrum of sugar compounds in the fasting state and during an oral glucose tolerance test (OGTT) in healthy, prediabetic, and type 2 diabetic volunteers.
Methods and results: The three volunteer groups underwent a standard OGTT. Plasma samples obtained in the fasting state, 30 and 90 min after the OGTT, are subjected to a semitargeted GC–MS (gas chromatography–mass spectrometry) sugar profiling. Overall, 40 sugars are detected in plasma, of which some are yet unknown to change during an OGTT. Several sugars (e.g., trehalose) reveal significant differences between the volunteer groups both in fasting plasma and in distinct time courses after the OGTT. This suggests an endogenous production from orally absorbed glucose and/or an insulin-dependent production/removal from plasma.
Conclusion: It is demonstrated that more sugars than expected can be found in human plasma. Since some of these show characteristic differences depending on health status, it may be worthwhile to assess their usability as biomarkers for diagnosing early-stage insulin resistance and type 2 diabetes.

1. Introduction

Although glucose and insulin, or HbA1c levels are routinely determined for type 2 diabetes management, they do not reveal which metabolic perturbations occur during glucose consumption, as simulated by an oral glucose tolerance test (OGTT).[1–5] Therefore, metabolomics has become popular as a more comprehensive profiling approach and biomarker discovery tool. This approach has revealed numerous differences in plasma or urine metabolites between healthy subjects and subjects with impaired glucose tolerance and/or type 2 diabetes.[4,6–12] Some studies have not only searched for differences in metabolite profiles during fasting, but also in time courses during an OGTT.[1,3,13–18] Remarkably, these studies mainly identified plasma amino acids, various lipid species, or bile acids as discriminating markers related to insulin resistance (IR) and type 2 diabetes. Considering diabetes is associated with marked alterations in carbohydrate metabolism, it is thus surprising that other sugars or sugar-derived intermediates have rarely been identified as associated with the metabolic impairments.

We recently established a semitargeted sugar profiling method and demonstrated a rather complex mixture comprising 55 different sugar compounds in urine samples of healthy human volunteers.[19] This finding suggested that the plasma sugar profile is equally complex. Metabolite profiling studies of human biofluids typically analyze only a few sugar compounds; this is mainly due to analytical limitations such as insufficient separation and identification of structural isomers.[19–22] Therefore, sugars or sugar-like candidates are

C. I. Mack, Dr. C. H. Weinert, B. Egert, Prof. S. E. Kulling
Department of Safety and Quality of Fruit and Vegetables
Max Rubner-Institut
Haid-und-Neu-Strasse 9, Karlsruhe 76131, Germany
E-mail: Carina.Mack@mri.bund.de

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/mnfr.201901190

DOI: 10.1002/mnfr.201901190

Dr. P. G. Ferrario
Department of Physiology and Biochemistry of Nutrition
Max Rubner-Institut
Haid-und-Neu-Strasse 9, Karlsruhe 76131, Germany
Dr. A. S. Hoefle, Y.-M. Lee, Prof. H. Daniel
Department of Food and Nutrition
Technical University of Munich
Gregor-Mendel-Strasse 2, Freising-Weihenstephan 85354, Germany
Prof. T. Skurk
Core Facility Human Studies
ZIEL Institute for Food and Health, Technical University of Munich
Gregor-Mendel-Strasse 2, Freising-Weihenstephan 85354, Germany
Prof. T. Skurk
Else Kröner-Fresenius-Center for Nutritional Medicine
Technical University of Munich
Gregor-Mendel-Strasse 2, Freising-Weihenstephan 85354, Germany

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

Molecular Nutrition & Food Research
www.mnf-journal.com

Table 1. Basic characteristics of the three volunteer groups.

	Healthy			Prediabetic			Diabetic		
	All	Male	Female	All	Male	Female	All	Male	Female
n	—	15	—	15	10	5	11	3	8
Age [years][a)]	—	26.3 ± 2.5	—	62.1 ± 6.5	60.0 ± 5.9	66.2 ± 6.1	63.0 ± 6.2	65.3 ± 4.7	61.9 ± 6.6
BMI [kg cm^{-2}][a)]	—	24.0 ± 2.0	—	29.0 ± 5.9	28.5± 6.2	30.1 ± 5.6	29.4 ± 2.6	28.9 ± 1.8	29.5 ± 2.9

[a)] For age and BMI mean values ± standard deviations are reported.

often reported as unknowns,[8,14] or are represented as a group, for example "hexoses"[7,9,14] or "fructose/glucose/galactose,"[13,18] which clearly limits interpretability. In addition to glucose, sugar compounds identified as changing during the time course of an OGTT or that display altered levels in type 2 diabetes patients are 1,5-anhydroglucitol, mannose, maltose, and glucuronic acid.[11,12,14,15,20,23] These findings suggest that additional sugars, and intermediates of the metabolic conversion of glucose, may be found in plasma and urine when specific methods are employed.

We therefore applied our semitargeted sugar profiling method, previously used for urine analysis,[19] to assess whether novel sugar compounds can be identified in plasma of healthy, prediabetic, and diabetic subjects in a fasted state and after an OGTT.

2. Experimental Section

2.1. Study Design and Subjects

This pilot study was part of the MIPROMET cohort, which was performed at the Human Study Center of the Technical University of Munich at the Weihenstephan campus.[24] Study design, participant recruitment, and examination procedures for healthy and prediabetic participants were previously described by Hoefle et al.[24] Type 2 diabetic participants were additionally recruited with study design and research parameters identical to those of the healthy and prediabetic volunteers.[24] For patients with type 2 diabetes, oral medication (metformin alone) and/or insulin therapy was an additional inclusion criterion. Prediabetic and diabetic states were defined according to WHO criteria.[25] Altogether, 15 male healthy controls, 15 male and female prediabetic cases, and 11 male and female diabetic cases were recruited for this pilot study. However, distribution of males and females, as well as age class was unbalanced between volunteer groups (see **Table 1** for volunteer characteristics) because of difficulties with recruitment due to strict inclusion criteria (see flow chart in Figure S1, Supporting Information File 1).

Volunteers received a test drink containing 50 g of maltodextrin19 (starch hydrolysate; Berco Arzneimittel, Kleve, Germany), 10 g of lactulose (Hemopharm, Bad Vilbel, Germany), 2 g of 4-hydroxyproline (Sigma-Aldrich, Munich, Germany), or 1 g of acetaminophen (Paracetamol; Ratiopharm, Ulm, Germany) in the case of prediabetic and diabetic volunteers, and ten drops of lemon or vanilla flavor (Dr. Oetker, Bielefeld, Germany). Inclusion of non-digestible lactulose enabled the assessment of the oro-cecal transit time by detecting hydrogen exhalation in the breath as a result of fermentative production of hydrogen by the intestinal flora. The appearance of acetaminophen in plasma depended on the velocity of gastric emptying as the limiting step, allowing the rate of appearance of paracetamol in the blood to serve as marker for gastric emptying[26]; similarly 4-hydroxyproline can be applied. Maltodextrin19 rather than D-glucose was used since this prevents secondary effects of high osmolarity on gastric emptying. Maltodextrin19 could be considered equivalent to glucose due to its fast hydrolysis in the gut. With a dextrose equivalent of 19, maltodextrin19 was close to the typically used glucose syrup in commercial OGTT solutions.

Plasma samples were taken in fasted state, as well as 30 and 90 min after ingestion of the test drink. Blood was collected in monovettes coated with EDTA, then inverted gently ten times and centrifuged at 20 °C at 1000 × *g* for 10 min. Thereafter, plasma was aliquoted on ice, frozen on dry ice, and stored at −80 °C until analysis.

All procedures performed in the study were in accordance with the 1964 Helsinki declaration and its later amendments. The study was approved by the Ethics Committee of the Technical University of Munich (#2436/09) and was registered at the German Clinical Study Register (DRKS00005682). Informed consent was obtained from all individual volunteers included in the study.

2.2. Semitargeted GC-MS Sugar Profiling

For plasma sample preparation, 20 µL of the internal standard solution (see Table S1, Supporting Information File 1) were pipetted into a 1.5 mL Eppendorf reaction tube before adding 40 µL plasma sample. Then, 150 µL of ice-cold methanol were added and the tube mixed for protein precipitation. After centrifugation for 10 min at 16 100 × *g* and 4 °C, 120 µL of the supernatant were transferred to a second 1.5 mL Eppendorf reaction tube. For lipid extraction, 700 µL tert-butyl methyl ether (TBME) were added and samples vigorously mixed for 2 min. After the addition of 70 µL of water, samples were again vigorously mixed for 2 min and then shortly centrifuged for phase separation. The upper TBME-containing phase was discarded and the aqueous phase transferred to a GC screw-top vial with a 300 µL insert. After rinsing the reaction tube with 30 µL methanol, the wash solution was added to the aqueous phase in the GC vial. Samples were then evaporated for 90 min at 40 °C and <1 mbar. For removal of traces of water, 20 µL of methanol were added and samples evaporated again under the same conditions for 30 min. Afterward, samples were methoximated using 20 µL methoxylaminhydrochloride in pyridine (20 mg mL^{-1}), the reaction was carried out for

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

Molecular Nutrition & Food Research
www.mnf-journal.com

6 h at 60 °C and 1000 rpm. Trimethylsilylation was carried out by adding 40 µL *N*-(trimethylsilyl)-*N*-methyltrifluoroacetamide (MSTFA) with 1% trimethyl chlorosilane (TMCS) and incubation for 1 h at 75 °C without shaking. Analyses were performed on a Shimadzu GCMS QP2010 Ultra instrument (see Table S2, Supporting Information File 1) in scheduled Scan-/SIM (selected ion monitoring)-mode allowing the selective and sensitive detection of sugar compounds. Measurements were performed with helium as carrier gas in constant pressure mode on a 60 m Rxi-5SilMS column with 10 m integrated pre-column, 0.25 mm internal diameter, and 0.25 µm film thickness. 1.2 µL sample were injected in cold split mode and a split ratio of 1:5. For further details about instrument parameters, see Table S3, Supporting Information File 1. The GCMSsolution (version 4.1.4, Postrun Analysis) software was used to pre-process the data (details see Table S4, Supporting Information File 1). For each measurement day, along with the study samples, quality control (QC) samples (mixture of all study samples) were regularly measured to enable correction of drift and offset effects and a solvent blank was created using water instead of the sample.[27] A retention index marker solution was added to the solvent blank.[27] Each day, 24 study samples, seven QC samples, and a solvent blank were prepared and analyzed, beginning with the daily blank, followed by four blocks of six study samples surrounded by QC samples, and at the beginning and end of the day two QC samples were always measured. Each measurement week, the liner was changed and the MS was tuned. The injector septum was replaced every 100 runs.

2.3. Analysis of the Ingested Test Drink for the OGTT and Further Commercial OGTT Solutions

A 2 mL sample of the ingested test drink for the OGTT was evaporated and then 10 mg of the dried powder were solved in 500 µL of deionized water. Afterward, 50, 100, and 200 µL of the solution were evaporated for GC-MS analysis. Derivatization and GC-MS analysis were the same as described earlier for plasma samples (see also Methods Section 1.2.1, Supporting Information File 1).

An additional GC × GC-MS measurement was performed to explore the sugar profile of the ingested test drink and two commercial OGTT solutions (Accu-Chek Dextrose O.G-T., Roche and Rapilose OGTT solution, Galen) in more detail. For the ingested test drink 25 µL of the solution described earlier were evaporated, while in the case of both commercial OGTT solutions 50 µL of a 1:20 dilution were evaporated. Derivatization was as described earlier for plasma samples (see also Methods Section 1.2.1, Supporting Information File 1). The GC × GC-MS parameters can be found in Table S6, Supporting Information File 1.

2.4. Statistics

2.4.1. Test for Group and Time Differences After an OGTT

The pilot study consisted of repeated measurements of sugar compounds at non-equidistant time points (0, 30, and 90 min) in three independent groups (15 healthy, 15 prediabetic, and 11 diabetic participants). Due to the participant recruitment, the pilot study had unbalanced participant characteristics. Participant recruitment reflected the fact that older people were more prone to develop type 2 diabetes mellitus; the average age of healthy participants was ≈26 years in contrast to 63 years in diabetic participants (Table 1). Moreover, the distribution of sex in the three groups was uneven (Table 1). A general model accounting for the characteristics leading to the unbalanced design would describe each sugar compound dependent on: i) group, time, and interaction between both, ii) age, BMI, and sex (as competing covariates), and iii) with random slope and random intercept to account for correlations across time (longitudinal data). An example of such possible models can be found in File 2, Supporting Information.

However, according to the one-in-ten rule, it is not reasonable to fit such a general model with the limited sample size in this pilot study. To find a compromise between model complexity and feasibility, a simpler model (mixed linear model) was fitted where each sugar compound was described dependent on group and time with a random intercept (to account for correlation between repeated measurements at different time points for each participant). Testing was performed for group and time differences in order to identify and focus on sugar compounds, which change concentration differently in healthy, prediabetic, and diabetic participants after an OGTT. Specifically, the differences in sugar compounds were tested for i) healthy versus diabetic group, ii) healthy versus prediabetic group, and iii) prediabetic versus diabetic group, while keeping time constant. Moreover, the differences in sugar compounds were tested for i) 0 versus 30 min, ii) 0 versus 90 min, and iii) 30 versus 90 min, while keeping group constant (see File 3, Supporting Information).

For the sake of simplicity and to ensure comparability, normality for sugar compound variables was generally assumed, although literature has shown that this assumption is not always justified.[28,29] For model diagnostics, QQ-plots depicting the quantiles of the residuals against the theoretical normal were included to ascertain normality (File 3, Supporting Information). For those sugar compounds found to be significantly different either dependent on health status and/or on time, a visual analysis of these QQ-plots was performed. For the majority of sugar compounds, the assumption of normality was justified.

The software R (version 3.3.2 2016-10-31)[30] was used to perform these statistical analyses. The applied R packages can be found in[31–37] and were added as libraries in Files 2 and 3, Supporting Information. A detailed report (including the R script) of the main statistical analyses using R can be found in File 3, Supporting Information. Generally, when no value was reported due to the signal intensity being below the limit of detection, half of the minimal signal intensity for this sugar compound was considered.

2.4.2. Differentiation of Healthy, Prediabetic, and Diabetic Subjects in the Fasted State

Independently of the administration of maltodextrin19, some sugar compounds already differ between the healthy, prediabetic, and diabetic group in the fasted state. A principal component

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

Molecular Nutrition & Food Research
www.mnf-journal.com

analysis (PCA) and a one-way ANOVA were performed in order to identify such sugar compounds.

First, the PCA was performed to observe whether plasma sugar profiles in the fasted state enable a separation of healthy, prediabetic, and diabetic volunteers. Both the scores and the loadings plot were drawn. The loadings plot was evaluated to ascertain whether glucose was the main variable driving the separation of the three groups.

Second, to determine the sugar compounds differentiating between the three groups, an ANOVA-based response screening workflow was applied. Briefly, a first one-way ANOVA with correction for multiple testing was performed to screen for significant sugar compounds, which were then tested for normality and homogeneity of variances. A second one-way ANOVA and post-hoc tests (each depending on the distribution and homogeneity of variance) were performed to receive the final selection of sugar compounds significantly differentiating between the healthy, prediabetic, and diabetic group in the fasted state. Specifically, the second one-way ANOVA was a normal one-way ANOVA (in the case of normality and homogeneity of variances), a one-way ANOVA with Welch-correction (in the case of normality and inhomogeneity of variances), or a one-way ANOVA with Kruskal–Wallis-correction (in the case of non-normality). For a more detailed description of the screening workflow, the reader is referred to the Supporting Information by Weinert et al.[38] Box plots were plotted for the significant sugar compounds revealed by these procedures. The software JMP (version 13, SAS Institute Inc., Cary, NC, 1989–2007) was used to perform these statistical analyses.

2.5. Evaluation of Potential Confounders in Fasting Plasma Samples in an Independent Study

In the present study, the healthy group comprised solely male volunteers, the prediabetic group was 2/3 male, and the diabetic group was 2/3 female (Table 1). Moreover, in addition to sex, also age and body mass may represent important baseline covariates (Table 1). Due to the sample size limitation of this pilot study, another independent external study was considered to generally ascertain the influence of age, BMI, and sex on the sugar compounds of interest using correlation analyses and unpaired tests.

The sugar profile of a subset of fasting plasma samples from the cross-sectional KarMeN study was analyzed. Details about study design, participant recruitment, and examination procedures, as well as sample storage were described by Bub et al.[39] Sample preparation, data preprocessing, and treatment were exactly the same as described in Section 2.2 and in Section 1.2.1, Supporting Information File 1. Minor changes in the instrument parameters can be found in Table S3, Supporting Information File 1. For participant characteristics of the randomly selected subset of 58 fasting plasma samples (about ten participants for every 10 years with half of the participants being male) see Table S5, Supporting Information File 1.

In particular, to ascertain the influence of age, BMI, and sex on sugar compounds found to differentiate between healthy, prediabetic, and diabetic subjects, a correlation analysis (Spearman correlation coefficients), *t*-tests (in the case of normality and homogeneity of variances), Welch's *t*-tests (in the case of normality and inhomogeneity of variances), or Wilcoxon *t*-tests (in the case of non-normality) were performed. In the case of age, two groups of the KarMeN volunteers were defined covering the same age range as the healthy (21–30 years; $n = 10$) and prediabetic and diabetic volunteers (50–72 years; $n = 23$) of the MIPROMET study. For BMI, KarMeN participants were divided into two groups (19–23.9 kg m^{-2}, $n = 29$ and 24–30 kg m^{-2}, $n = 29$). In the case of sex, 28 males were compared with 30 females. The KarMeN study was registered at the German Clinical Study Register (DRKS00004890) and was in accordance with the 1964 Helsinki declaration and its later amendments.

The software JMP (version 13, SAS Institute Inc., Cary, NC, 1989–2007) was used to perform these statistical analyses.

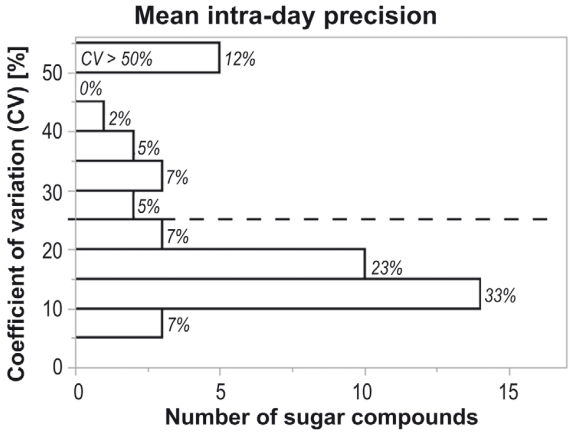


Figure 1. Distribution of the average intra-day coefficients of variation for all sugar compounds measured in the QC samples before drift correction. Dashed line indicates coefficients of variation below and above 25%.

3. Results

3.1. Plasma Sugar Profile

Overall, we detected 40 sugar compounds represented by mono- and disaccharides, polyols, sugar acids, and so far unknown sugar-like compounds with the semitargeted sugar profiling method. Based on characteristic masses in the scan mass spectra, unknown compounds were defined as sugar-like compounds (e.g., m/z 292 and 333 are characteristic for sugar acids, m/z 319 and 307 for aldoses, ketoses as well as polyols, m/z 305 and 318 for cyclitols, and m/z 361 is characteristic for disaccharides). Additionally, in some cases the chromatographic elution order allowed us to derive the number of C-atoms in the sugar compound (e.g., pentose or hexose). However, future absolute identification can only be ensured by comparing retention indices and mass spectra of the unknown sugar-like compounds (Figure S2, Supporting Information File 1) with those of known standard substances. Data on the quality of measurement (e.g., reproducibility of quality control samples and internal standards, identification level, median, and range from minimum to maximum of peak area in the study samples) and on selectivity can be found in Table S7, Supporting Information File 1. **Figure 1**

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

Molecular Nutrition & Food Research
www.mnf-journal.com

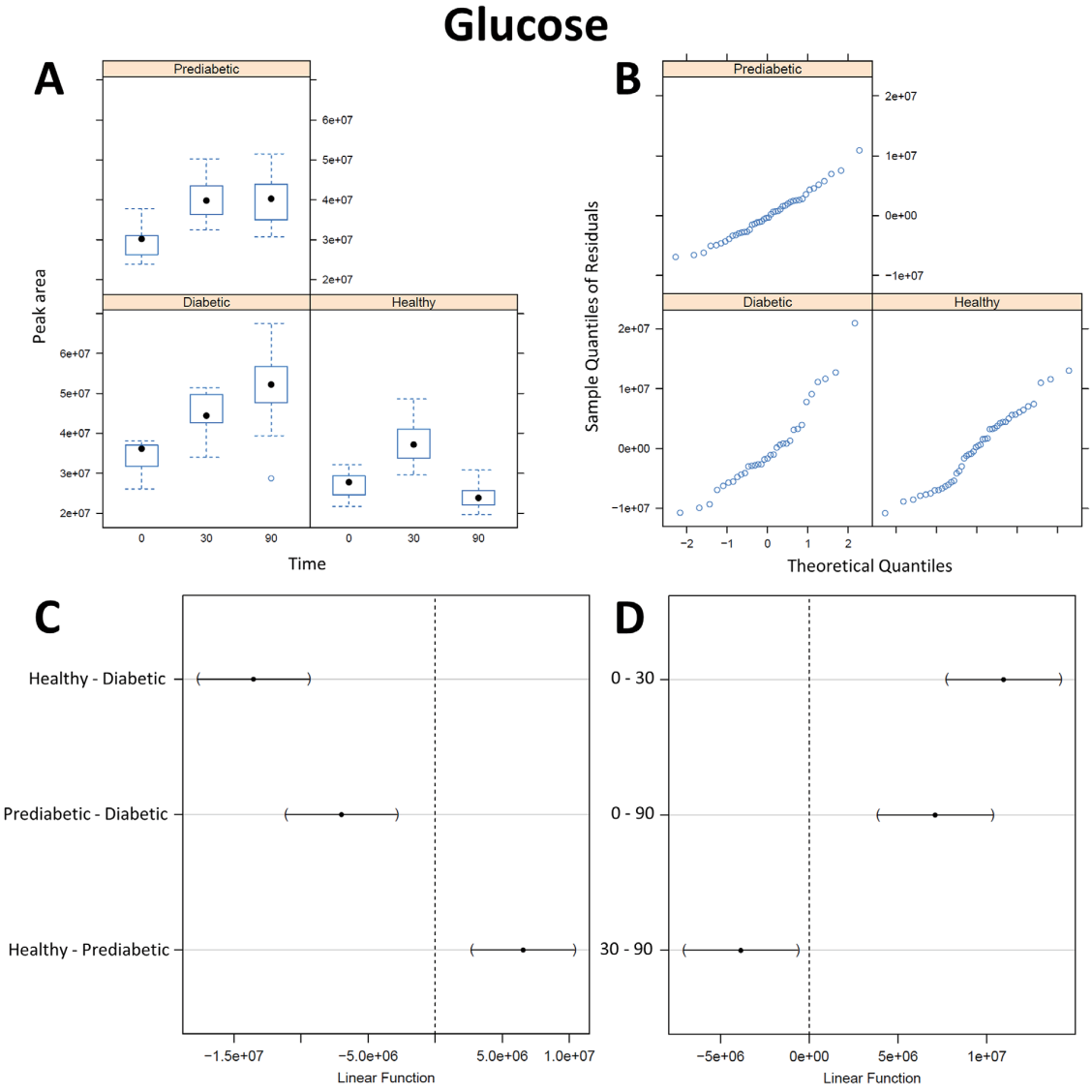


Figure 2. Glucose as example for the statistical workflow to detect group and time differences after an OGTT. Panel A) Box plots grouped by health status and time. Panel B) QQ-plots depicting the quantiles of the residuals against the theoretical normal. Panel C) Plot for the effect (dots) and confidence interval (lines) for differences in the health status (group). Panel D) Plot for the effect (dots) and confidence interval (line) for differences during the time points (time). Significance corresponds to the confidence interval line not crossing the vertical dotted zero-line.

depicts the distribution of the average intra-day coefficients of variation for all sugar compounds measured in the QC samples before drift correction. 70% of the sugar compounds had a coefficient of variation of less than 25% and could therefore be measured with good reproducibly. Interestingly, disaccharides such as sucrose, lactose, maltose, trehalose, and lactulose were detected in all plasma samples, even though carbohydrates are generally hydrolyzed into monosaccharides before absorption. Lactulose was only detected in the time course after ingestion of the test drink containing lactulose for transit-time measurement (Figure S3, Supporting Information File 1).

3.2. Results for Group and Time Differences After an OGTT

The longitudinal trend of each sugar compound was illustrated by box plots grouped by health status and time points. Each sugar compound was modeled by a mixed linear model on group and time (both as factor variables) with a random intercept. Then, we tested for all possible group and time differences. In Supporting Information File 3, each sugar compound is presented by box plots, the fitted model, test results, and QQ-plots (see e.g., glucose in **Figure 2**). **Table 2** summarizes the sugar compounds showing significant group and/or time differences. Moreover, in Table 2,

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

Molecular Nutrition & Food Research
www.mnf-journal.com

Table 2. Significant sugar compounds to evaluate health status-dependent changes in time courses during an OGTT.

Sugar compounds	*p*-values—Group			*p*-values—Time			Direction of time-dependent change[a)]		
	Healthy—Diabetic	Prediabetic—Diabetic	Prediabetic—Healthy	0–30 min	0–90 min	30–90 min	Healthy	Prediabetic	Diabetic
Glucose[b),d)]	<0.001	<0.001	<0.001	<0.001	<0.001	0.013	↑↓	↑	↑
Mannose[c)]	<0.001		<0.001			<0.001	↓	–	↑
1,5-Anhydroglucitol	<0.001		<0.001	0.044			↑↓	–	–
U15	<0.001		0.004	<0.001	0.041		↑↓	↑	↑
Sedoheptulose[b),d)]	0.001		0.03				↑↓	–	–
Fructose[b),d)]	<0.001	<0.001		0.005	0.002		↑↓	↑	↑
Trehalose[c)]	<0.001	<0.001		<0.001	<0.001	0.004	↑↓	↑	↑
Sorbitol[b)]	<0.001	<0.001					↑↓	↑	↑
Arabinose[b)]	<0.001			0.014			↑↓	↑	↑
U01	0.001				0.035		–	↓	↓
U07	0.007			<0.001	0.013		↑↓	↑	↑
U16	0.004			0.018			–	↑↓	↑
meso-Erythritol	0.038						↑↓	–	–
Glucuronic acid[d)]		0.027					↓	–	–
Psicose				<0.001	<0.001	<0.001	↑	↑	↑
U18[b)]	0.015	0.012		<0.001	<0.001	<0.001	↑	↑	↑
Threitol				<0.001	<0.001		↑	↑	–
Lactulose[b),d)]				0.009	<0.001	<0.001	↑	↑	↑
scyllo-Inositol				0.039	0.002		–	↑	↑
Xylose[b)]	0.049				<0.001	<0.001	↓	–	–
Erythronic acid				0.036			↑↓	↑	↑
Maltose[b),d)]					0.032		↑↓	↑	↑
U17					0.034		↓	–	↓
Threonic acid						0.002	↑↓	↑	–

Significance threshold: *p*-value < 0.05. [a)] The direction of the time-dependent changes for the different volunteer groups is given, ↑↓: increase from 0 to 30 min and decrease from 30 to 90 min; ↑ increase from 0 to 30 to 90 min; ↓ decrease from 0 to 30 to 90 min;—remain unchanged over time; [b)] Detectable in the OGTT test drink; [c)] Presence in OGTT drink not confirmed due to low concentration and huge amounts of glucose or maltose; [d)] For readability only one of two detected derivatives is listed.

the direction of time-dependent changes of sugar compounds is described. Box plots of these sugar compounds are depicted in **Figure 3**.

Specifically, the selected sugar compounds could be classified depending on whether they mainly show a health status-dependent difference, either dependent or independent of the time course during an OGTT, or they mainly show a difference in time course independent of health status. For example, i) trehalose clearly differentiated between health status and also between the time points (group and time effect), ii) sorbitol showed a distinct difference only in health status (only group effect), and iii) psicose distinctly increased during the OGTT independent of the health status (only time effect). For the effects of all other sugar compounds see Table 2.

3.3. Differentiation of Healthy, Prediabetic, and Diabetic Subjects in the Fasted State

To ascertain whether healthy, prediabetic, and diabetic volunteers can be separated based on their fasting plasma sugar profile, we compared PCA scores A) and loadings B) plots (**Figure 4**). A separation of healthy and diabetic subjects was observed in the scores plot, while the prediabetic volunteers were in between. In the loadings plot it is obvious that, in addition to glucose, a range of other sugar compounds similarly added to the separation of the volunteer groups observed in the scores plot.

In addition to glucose, significant differences between healthy, prediabetic, and diabetic volunteers were revealed for maltose, trehalose, an unknown sugar compound (U15), fructose, mannose, 1,5-anhydroglucitol, and sedoheptulose (**Figure 5**). The highest median fold change with a factor of almost three between healthy and diabetic volunteers (diabetic volunteers had a higher level) was found for maltose. Sedoheptulose concentration was almost twofold lower in diabetic volunteers than in healthy individuals.

3.4. Evaluation of Potential Confounders in Fasting Plasma Samples in an Independent Study

To assess potential confounder effects of age, BMI, and sex across the volunteer groups in the set of sugars with significant

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

Molecular Nutrition & Food Research
www.mnf-journal.com

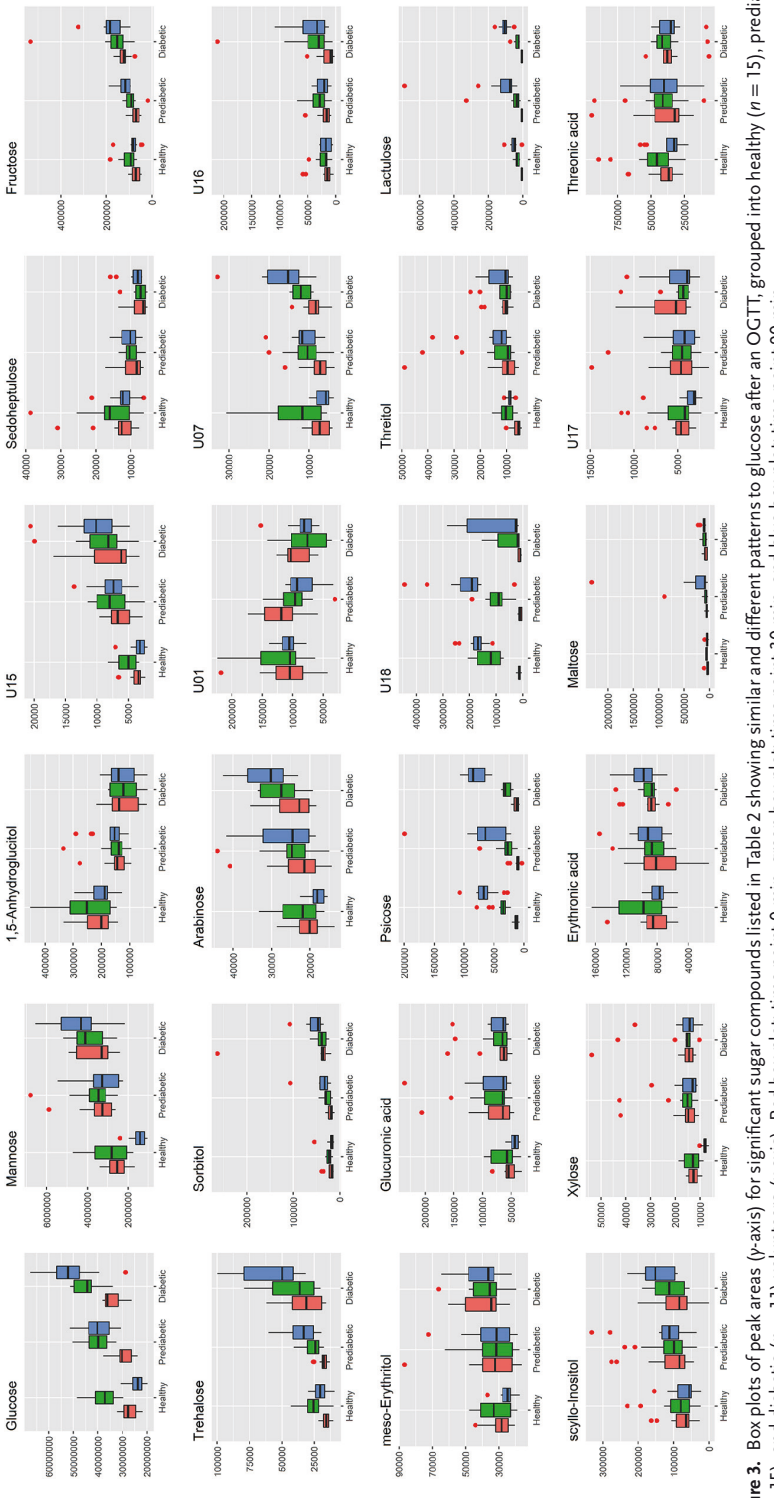


Figure 3. Box plots of peak areas (*y*-axis) for significant sugar compounds listed in Table 2 showing similar and different patterns to glucose after an OGTT, grouped into healthy ($n = 15$), prediabetic ($n = 15$), and diabetic ($n = 11$) volunteers (*x*-axis). Red box plot: time point 0 min, green box plot: time point 30 min, and blue box plot: time point 90 min.

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

Molecular Nutrition & Food Research
www.mnf-journal.com

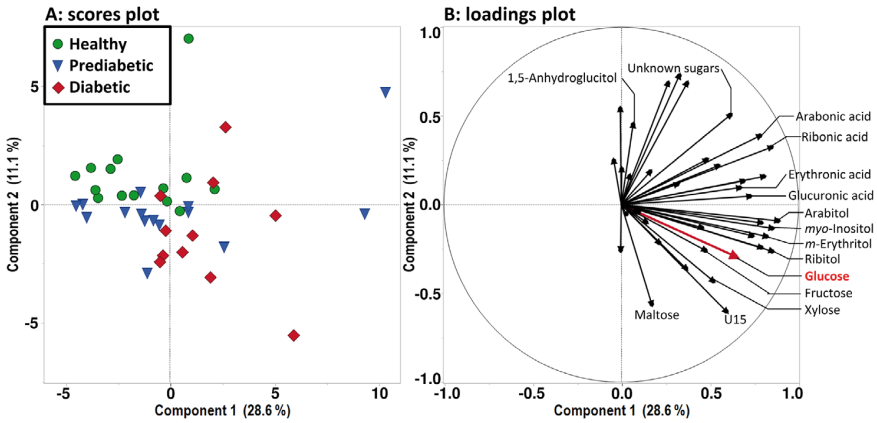


Figure 4. Separation of healthy ($n = 15$), prediabetic ($n = 15$), and diabetic ($n = 11$) volunteers based on fasting plasma sugar profile using a principal component analysis (PCA). A) scores plot; B) loadings plot; glucose marked as bold, red arrow.

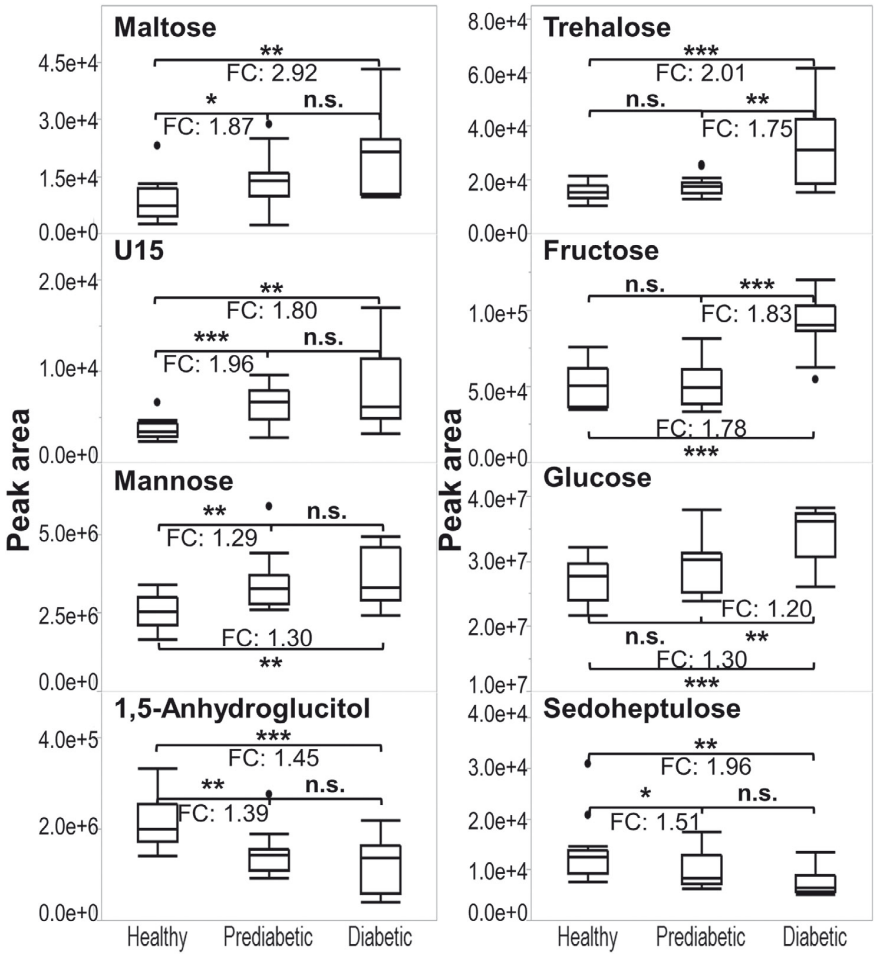


Figure 5. Box plots of peak areas for sugar compounds identified as significantly distinguishing between healthy ($n = 15$), prediabetic ($n = 15$), and diabetic ($n = 11$) volunteers based on fasting plasma using the response screening workflow. *$p \leq 0.05$; **$p \leq 0.01$; and ***$p \leq 0.001$; n.s., not significant. FC, median fold change. Only one of the two detected derivatives is shown in the case of maltose, fructose, glucose, and sedoheptulose. In the case of glucose and 1,5-anhydroglucitol, the 2nd ANOVA was a normal one-way ANOVA and the post-hoc test was a Tukey-HSD, while all other sugar compounds were subjected to a Wilcoxon test as the 2nd ANOVA and a Steel–Dwass as post-hoc test.

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

Molecular Nutrition & Food Research
www.mnf-journal.com

Table 3. Results of the correlation analysis and unpaired hypothesis tests for verification of potential relevant sugar compounds using fasting plasma of a selected subgroup of the KarMeN cohort.

Sugar compounds	Correlation				*t*-test[a]					
	Age		BMI		Age		BMI		Sex	
	Spearman's rho	*p*-value	Spearman's rho	p-value	*p*-value	Median fold change	*p*-value	Median fold change	*p*-value	Median fold change
Threonic acid	−0.286	0.029	−0.297	0.024	**0.008**	**−1.25**	0.023[c]	−1.15	0.283	−1.07
Sedoheptulose[b]	−0.229	0.084	−0.144	0.282	**<0.001**	**−1.57**	0.417	−1.08	**0.018**	**1.33**
U07	−0.102	0.448	−0.066	0.624	0.232	−1.08	0.303	1.01	0.269	−1.02
1,5-Anhydroglucitol	−0.074	0.580	0.070	0.601	0.604	1.03	0.546	1.02	**<0.001**[c]	**1.42**
Trehalose	−0.038	0.778	0.238	0.072	0.605	−1.11	0.634	1.14	0.077	1.25
scyllo-Inositol	−0.002	0.987	−0.172	0.196	0.430	−1.17	0.133	−1.34	0.831	1.04
U18	−0.001	0.992	−0.044	0.744	0.572	−1.02	0.836	−1.00	0.827[2]	1.06
Sorbitol	0.043	0.746	0.026	0.849	0.660	−1.02	0.517	−1.05	0.686	−1.10
Maltose[b]	0.055	0.682	0.173	0.193	0.976	−1.05	0.150	1.09	0.944	−1.04
Fructose[b),d]	0.058	0.667	0.092	0.492	0.433	1.03	0.549	1.06	0.926	1.01
U16	0.075	0.578	**0.315**	**0.016**	0.171	−1.05	0.170	1.13	0.054	1.10
Glucose[b]	0.121	0.368	0.128	0.340	0.572	1.01	0.290	−1.00	0.233	1.01
U01	0.188	0.158	0.120	0.369	0.812	1.01	0.917	1.02	0.133	−1.11
Ribitol[d]	0.245	0.064	−0.068	0.610	0.060	1.12	0.247	−1.03	0.017	−1.14
U15[d]	0.252	0.057	0.264	0.045	0.137	1.32	0.164	1.40	0.343	1.26
Psicose[d]	0.267	0.043	0.270	0.041	0.389	1.02	0.130	1.15	0.016	1.15
Arabinose	0.292	0.026	0.283	0.031	0.226	1.07	0.808	1.09	0.349	1.05
Mannose	0.297	0.024	0.161	0.229	0.086	1.17	0.178	1.12	0.050	1.04
Xylose[d]	**0.327**	**0.012**	0.246	0.063	0.327	1.04	0.101	1.05	0.437	1.03
U17	**0.374**	**0.004**	0.099	0.461	0.123	1.24	0.701	1.06	0.891	1.00
Threitol[d]	**0.415**	**0.001**	**0.302**	**0.021**	0.078	1.10	0.070	1.08	0.828	−1.05
Erythronic acid	**0.441**	**<0.001**	**0.467**	**<0.001**	0.239	−1.00	0.002	1.14	0.185	1.05
meso-Erythritol[d]	**0.477**	**<0.001**	**0.443**	**<0.001**	0.021	1.11	0.012	1.16	0.815	−1.01
Glucuronic acid[b]	**0.547**	**<0.001**	**0.418**	**0.001**	0.059	1.18	0.016	1.13	0.058	1.08

[a] For the *t*-test the following groups were defined for comparison: age: 21–30 years ($n = 10$) against 50–72 years ($n = 23$); BMI: 19–23.9 kg m^{-2} against 24–30 kg m^{-2} (each $n = 29$) and sex: male ($n = 28$) against female ($n = 30$). Significant values ($p < 0.05$) with rho > $|0.3|$ or with median fold changes higher $|1.25|$ are highlighted in bold; [b] For readability only one of two detected derivatives is listed; [c] In these cases a Welch's *t*-test was applied due to unequal variances; [d] *t*-test was a non-parametric Wilcoxon test, while for the other sugar compounds normal one-way ANOVA was used (depending on normality and homogeneity of variances).

differences between groups, fasting plasma samples selected from the KarMeN cohort were analyzed for comparison. Spearman correlation coefficients as well as results for hypothesis tests with variables such as age, BMI, and sex covering ranges similar to those of the volunteers in the present study are shown in **Table 3**. They revealed a significant correlation with age for xylose and the unknown sugar U17, in the case of threitol, erythronic acid, *meso*-erythritol, and glucuronic acid for both age and BMI, and in the case of unknown sugar U16 for BMI (Figure S4, Supporting Information File 1). Significant differences in the hypothesis testing were found for threonic acid regarding age, sedoheptulose regarding age and sex, and for 1,5-anhydroglucitol regarding sex (Figures S5 and S6, Supporting Information File 1). Thus, for some sugar compounds (xylose, threonic acid, threitol, *meso*-erythritol, glucuronic acid, sedoheptulose, and 1,5-anhydroglucitol) an extended general model accounting for the confounders, as suggested in Supporting Information File 2, would be more suitable and especially should be taken into consideration for subsequent studies. Since all other sugar compounds were not significantly affected by these variables (age, BMI, and sex) or had weak correlations (rho < $|0.3|$) or median fold changes below $|1.25|$, it is likely that these variables (age, BMI, and sex) are not confounding results for these sugar compounds and the simpler model applied in this pilot study (Supporting Information File 3) should be sufficient.

3.5. Analysis of the Ingested OGTT Test Drink and Further Commercial OGTT Solutions

To ascertain whether some of the significant sugar compounds listed in Table 2 might be occurring or also increase in plasma due to their presence in the test drink used for the OGTT, this test drink was analyzed using the same method as for the plasma sugar profiling. U18, one of the two sugar compounds showing a distinct time-dependent increase after the OGTT (Figure 3) was detected in the test drink. Furthermore, we found large amounts of tagatose, an epimer of fructose. Other sugar compounds were also detectable in the test drink, but in comparatively small amounts relative to glucose and maltose (Table 2).

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

Molecular Nutrition & Food Research
www.mnf-journal.com

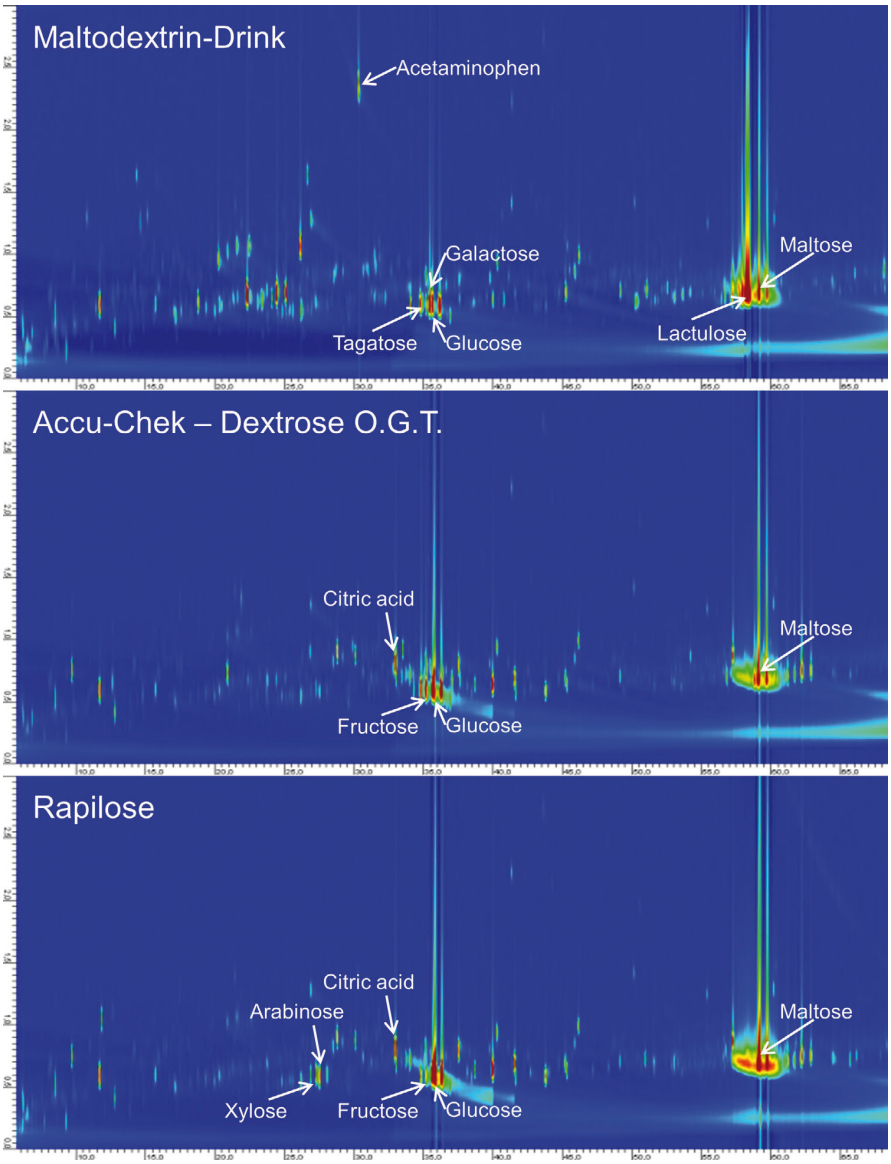


Figure 6. GC × GC-MS chromatograms from the start to 68 min of the maltodextrin-drink, and two commercial OGTT solutions. In the maltodextrin the additionally added acetaminophen and lactulose are highlighted. Some of the major differences between the OGTT solutions were also identified. The part of the chromatogram from 68 to 80 min was excluded due to many artifacts arising in this region.

Additionally, a general comparison of compounds occurring in the test drink applied in this study and further commercially available OGTT solutions was performed using GC × GC-MS measurements to assess whether the occurrence of other sugars and non-sugar compounds in OGTT solutions is common. Chromatograms of the GC × GC-MS measurements clearly show that the test drink and commercial OGTT solutions contained additional compounds to glucose, such as maltose, galactose, fructose, tagatose, and many more, even non-sugar compounds such as citric acid (**Figure 6**). In the case of the test drink from this study, one can also see the added acetaminophen and lactulose.

4. Discussion

4.1. Plasma Sugar Profile

Around 40 different sugar compounds, 27 known and 13 unknown, were reliably and reproducibly detected in plasma. Among the known mono- and disaccharides, polyols and sugar acids, we also detected a variety of rare sugar compounds such as psicose, trehalose, or xylonic acid. Interestingly, amounts of disaccharides such as sucrose, lactose, maltose, trehalose, and lactulose found in plasma were similar to other monosaccharides

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

Molecular Nutrition & Food Research
www.mnf-journal.com

(except glucose), challenging the dogma of complete hydrolysis of disaccharides by mucosal disaccharidases prior to absorption of monomers. Transfer of disaccharides into systemic circulation was already described decades ago,[40,41] but has recently received new attention since passive permeation of disaccharides such as lactulose across the epithelium can serve as a measure of gut permeability.[42] We here demonstrate the appearance of lactulose from the test drink (assessment of oral-cecal transit time) in peripheral blood. We also detected lactose, which in urine has been described as candidate marker for dairy consumption[19,43] and similarly, sucrose in urine was proposed as a marker for total dietary sugar intake.[44] The origin of the other detected disaccharides is currently a mystery. They may come from the diet, or they might be generated endogenously, for example by degradation of heteroglycans, or as by-product of other metabolic processes.

4.2. Sugar Compounds and Diabetes

The fasting plasma sugar profile already allowed a separation of healthy and diabetic volunteers, with prediabetic volunteers in between. The loadings plot revealed that other sugar compounds are just as relevant for this separation as the "key separator" glucose. Although glucose is a good and easily assessable clinical biomarker for both diabetic and prediabetic states, a metabolite profile may enable simpler detection of disease subtypes, for example to better predict complications or optimize preventive measures.[1] In addition to known marker compounds, such as 1,5-anhydroglucitol,[45] mannose,[46,47] and fructose,[12,48] other sugar compounds such as maltose,[11] trehalose, sedoheptulose, and an unknown sugar-like compound (U15) were found to significantly distinguish between healthy, prediabetic, and diabetic volunteers based on their fasting plasma levels. They thus represent promising biomarker candidates for IR and type 2 diabetes worth to be further explored in population studies. Such markers are important in the light of the ever-increasing incidence of diabetes mellitus type 2, as tools for diagnosing and preventing the disease as early as possible.[2,10,11,49]

1,5-anhydroglucitol is used as an indicator of short-term glycemic control in type 2 diabetes.[45] When the tubular load of glucose is high, 1,5-anhydroglucitol is rapidly excreted in urine due to competition for reabsorption via the renal brush-border membrane glucose transporters.[45] However, 1,5-anhydroglucitol plasma levels differed in males and females in the KarMeN cohort, suggesting that different threshold levels for diagnostic purposes may be needed.[50] Other sugar compounds such as xylose, threitol, sedoheptulose, *meso*-erythritol, erythronic acid, threonic acid, glucuronic acid, and two unknown sugars (U16, U17) were also associated with age, BMI, or sex in the KarMeN plasma sample set, and therefore, should be interpreted with caution when conducting subsequent studies. Recently, sedoheptulose was reported to decrease with age in urine sample data from the same cohort,[51] which suggests that differences found between healthy, prediabetic, and diabetic volunteers are mainly due to differences in age. The correlation of *meso*-erythritol with BMI is interesting with regard to a recent publication where erythritol was described as associated with weight gain in young adults.[52]

Other confounders could be the diet on the day before the intervention or more generally the habitual diets with respect to high or low sugar intake or consumption of sugar substitutes (by prediabetic and diabetic volunteers) such as sorbitol, mannitol, or erythritol. However, overnight fasting (12 h) minimizes the influence of the diet on the day before due to the fast clearance and excretion of the very polar sugar compounds. For those sugar compounds whose plasma levels changed during the OGTT, the effects of previous consumption can in essence be ruled out.

The disaccharide trehalose was found in higher levels in fasting plasma of diabetic volunteers, and plasma levels also changed differently during the OGTT. Although the changes mimic those of glucose, with a transient peak at 30 min and a decline at 90 min in healthy, and a steady increase over time in prediabetic and diabetic individuals, it cannot be confirmed that these changes are similarly affected by insulin and IR as known for glucose. To our knowledge, such effects on plasma trehalose have not been reported before, but an increased activity of plasma trehalase was described in diabetic states [53,54] and single nucleotide polymorphisms in the trehalase gene were recently found to be associated with type 2 diabetes.[55] Additionally, in a study on diabetic retinopathy, trehalose was detectable in most diabetic, but not in non-diabetic participants.[23] However, neither the origin of trehalose in plasma, nor the cause of the differences found between healthy and diabetic individuals are currently known.

Trehalose could not be detected in the test drink, and thus its appearance and time-profile in plasma argues for an endogenous origin. That maltose as a degradation product of maltodextrin provided in the test drink becomes detectable in plasma is in line with a significant permeation of intact disaccharides across the intestinal epithelium. What makes the findings on the disaccharides even more interesting is that their change over time in plasma during the OGTT is similar to the pattern of glucose, suggesting similar underlying insulin-dependent processes in clearance. This also holds true for mannose, fructose, xylose, sorbitol, *scyllo*-inositol, *meso*-erythritol, threitol, ribitol, arabinose, and two of the unknown sugar-like compounds (U07, U15). An endogenous production from glucose cannot be confirmed for these compounds as a result of the detection of trace quantities of fructose, sorbitol, xylose, sedoheptulose, and arabinose in the test drink itself. Nonetheless, the differences over time based on health status suggest underlying insulin-dependent processes and therefore these sugar compounds might be interesting for further exploration as biomarkers of IR and type 2 diabetes.

A comparison of the test drink applied in this study and further commercial OGTT solutions depicts clearly that further sugar and non-sugar compounds in addition to glucose can commonly be found in such test solutions. The additional sugar compounds likely arise from the commercial maltodextrin or glucose syrup used for production of OGTT solutions. This urges for care when interpreting postprandial changes of metabolites in plasma when using such test agents, as the occurrence of additional compounds might influence the metabolism of the compounds of interest and an endogenous production of additionally found compounds cannot be assumed. Therefore, a good characterization of the test products is an important step in any intervention study. More intensive work for the harmonization of the preparation of different OGTT solutions is desirable for comparable results in OGTT's.

Xylose has to our knowledge not been described as a metabolite that changes during the OGTT in volunteers with IR or diabetes. However, a recent analysis described different plasma levels in obese and lean individuals.[15] Xylose can be produced in the pentose phosphate pathway (PPP), which frequently shows impaired activity in diabetic states.[23] Although we detected sedoheptulose in plasma, further intermediates of the PPP such as ribulose and xylulose were only found in traces, and thus could not prove that PPP per se is altered. Similarly, arabinose could be a by-product of an epimerase reaction in PPP, but its production in human metabolism has not been confirmed.

Mannose is produced from glucose via fructose-6-phosphate, and has previously been identified as a diabetes metabolite marker.[11,20,56] Fasting mannose levels correlate with fasting plasma glucose, and mannose seems to be only minimally influenced by meals.[46] Early in vitro studies[57] demonstrated that mannose shares the uptake system of glucose in muscle, likely via Glut4, which is insulin-dependent. Consequently, mannose has a diagnostic marker quality, especially in the case of borderline diabetes,[46,47] and our data with the pronounced difference in mannose levels after the OGTT across the groups also demonstrates this special quality.

Two conspicuous sugar compounds, an unknown sugar-like compound U18 (possibly a C4 sugar acid) and psicose, showed a uniform strong increase after the OGTT, dependent on health status for U18 and independently of health status for psicose (U18 by 18.5-fold and psicose by sevenfold; Figure 3). Whereas U18 was detectable in the test drink, psicose was not, but we found high amounts of tagatose, which like psicose is an epimer of fructose. It may well be that tagatose was completely converted to psicose in human metabolism, explaining the steep postprandial increase in plasma psicose. Alternatively, psicose may also be produced endogenously from glucose via fructose in glycolysis by an epimerase in response to the glucose load. We also showed that fructose in plasma increases upon the glucose load, which requires that a fraction of the cytosolic fructose-phosphate pool is dephosphorylated, resulting in efflux of free fructose via a Glut-transporter system. Thus, it appears plausible that psicose as a fructose-epimer is generated from glucose in cells. Further research on the origin of psicose will be especially interesting in light of its proposed anti-hyperglycemic and anti-hyperlipidemic effects.[58–61]

Major limitations of our pilot study are the limited sample size, and the unbalanced design of the study, especially with respect to sex and age. A complex general modelling would have been able to adjust for these confounders. However, due to the limited sample size, a general model would result in overfitting. Therefore, to address the limitation of the unbalanced design as best as possible and generally assess the influence of age, BMI, and sex on sugar compounds of interest, an independent, external study was taken into account. Overall, our study can be viewed as a pilot study, which led to highly interesting and unexpected results about the influence of the diabetic state on the sugar profile after an OGTT. In subsequent studies with a larger sample size and a more balanced design, a general modelling accounting for confounders, interactions, random slope and intercept will be possible. Moreover, with a larger sample size, multivariate statistical analyses could additionally be conducted to focus on the different relationships between the sugar compounds.

5. Concluding Remarks

The 40 different sugar and sugar-like compounds we detected in human plasma were almost all recently identified by us as normal constituents of human urine.[19] This is an important finding with respect to large cohort studies, where often only plasma samples are available; meaning plasma sugar compounds are similar to 24 h urine samples, although concentrations in plasma are generally lower. In volunteers with IR or type 2 diabetes numerous sugar compounds show characteristic plasma profiles that mimic those of glucose, with rapid clearance during the OGTT in healthy individuals, but not in an IR or diabetic state. This suggests that, like glucose, many more sugar compounds are subject to insulin-dependent removal from plasma after absorption or when produced endogenously. Although we could not confirm that compounds such as psicose, trehalose, mannose, and others are indeed produced from glucose in metabolic pathways during the OGTT, for some this appears likely, suggesting an avenue for additional research, preferentially using stable-isotope labeled glucose. However, we demonstrated that the source of glucose for the OGTT, here in form of maltodextrin, can contain other sugars and additional compounds. We found this also holds true for commercial OGTT solutions, highlighting the need for good characterization of the test product and care in interpreting postprandial metabolite changes.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

The study presented here used biofluid material collected in the control arms of an intervention study (MIPROMET) on the role of whey and casein on metabolic responses (see ref. [24]), sponsored by Mondelēz International Inc. The present study was exclusively financed by the Max Rubner-Institut and the KarMeN study was financially supported by the German Federal Ministry of Food and Agriculture (BMEL). The work was part of the Foodball project (grant no. 2814ERA03E) funded by BMEL. The authors wish to thank Claudia Hoffmann, Silvia Remmert, and Michael Meyer of the GC-MS metabolomics laboratory at the Max Rubner-Institut for their excellent technical assistance and the staff of the Human Study Center at the Technical University of Munich for conducting the human intervention. The authors would also like to thank the staff of the Human Nutrition Study Center at the Max Rubner-Institut for their excellent work with the KarMeN study.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

S.E.K., H.D. designed the research project; H.D., T.S., Y.-M.L., A.S.H. developed the MIPROMET concept and design; C.I.M. conducted the analytical experiment; C.I.M., B.E., C.H.W., P.G.F. analyzed the data and performed statistical analysis; C.I.M. wrote the initial draft of the manuscript; C.H.W., S.E.K., H.D., P.G.F. critically reviewed and contributed to the manuscript;

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

Molecular Nutrition & Food Research
www.mnf-journal.com

C.I.M., H.D., S.E.K. had primary responsibility for final content. All authors read and approved the final version of the manuscript.

Keywords

gas chromatography, glucose, human plasma, mass spectrometry, sugar profiling, type 2 diabetes mellitus

Received: November 21, 2019
Revised: January 31, 2020
Published online: April 7, 2020

[1] O. Shaham, R. Wei, T. J. Wang, C. Ricciardi, G. D. Lewis, R. S. Vasan, S. A. Carr, R. Thadhani, R. E. Gerszten, V. K. Mootha, *Mol. Syst. Biol.* **2008**, *4*, 214.
[2] J. Chen, X. Zhao, J. Fritsche, P. Yin, P. Schmitt-Kopplin, W. Wang, X. Lu, H. U. Häring, E. D. Schleicher, R. Lehmann, G. Xu, *Anal. Chem.* **2008**, *80*, 1280.
[3] X. Zhao, A. Peter, J. Fritsche, M. Elcnerova, A. Fritsche, H. U. Haring, E. D. Schleicher, G. Xu, R. Lehmann, *Am. J. Physiol.-Endocrinol. Metab.* **2009**, *296*, E384.
[4] T. J. Wang, M. G. Larson, R. S. Vasan, S. Cheng, E. P. Rhee, E. McCabe, G. D. Lewis, C. S. Fox, P. F. Jacques, C. Fernandez, C. J. O'Donnell, S. A. Carr, V. K. Mootha, J. C. Florez, A. Souza, O. Melander, C. B. Clish, R. E. Gerszten, *Nat. Med.* **2011**, *17*, 448.
[5] J. R. Bain, R. D. Stevens, B. R. Wenner, O. Ilkayeva, D. M. Muoio, C. B. Newgard, *Diabetes* **2009**, *58*, 2429.
[6] X. Zhao, J. Fritsche, J. Wang, J. Chen, K. Rittig, P. Schmitt-Kopplin, A. Fritsche, H.-U. Häring, E. D. Schleicher, G. Xu, R. Lehmann, *Metabolomics* **2010**, *6*, 362.
[7] R. Wang-Sattler, Z. Yu, C. Herder, A. C. Messias, A. Floegel, Y. He, K. Heim, M. Campillos, C. Holzapfel, B. Thorand, H. Grallert, T. Xu, E. Bader, C. Huth, K. Mittelstrass, A. Doring, C. Meisinger, C. Gieger, C. Prehn, W. Roemisch-Margl, M. Carstensen, L. Xie, H. Yamanaka-Okumura, G. Xing, U. Ceglarek, J. Thiery, G. Giani, H. Lickert, X. Lin, Y. Li, et al., *Mol. Syst. Biol.* **2012**, *8*, 615.
[8] X. Li, Z. Xu, X. Lu, X. Yang, P. Yin, H. Kong, Y. Yu, G. Xu, *Anal. Chim. Acta* **2009**, *633*, 257.
[9] A. Floegel, N. Stefan, Z. Yu, K. Mühlenbruch, D. Drogan, H.-G. Joost, A. Fritsche, H.-U. Häring, M. Hrabě de Angelis, A. Peters, M. Roden, C. Prehn, R. Wang-Sattler, T. Illig, M. B. Schulze, J. Adamski, H. Boeing, T. Pischon, *Diabetes* **2013**, *62*, 639.
[10] X. Zhang, Y. Wang, F. Hao, X. Zhou, X. Han, H. Tang, L. Ji, *J. Proteome Res.* **2009**, *8*, 5188.
[11] K. Suhre, C. Meisinger, A. Döring, E. Altmaier, P. Belcredi, C. Gieger, D. Chang, M. V. Milburn, W. E. Gall, K. M. Weinberger, H.-W. Mewes, M. Hrabé de Angelis, H. E. Wichmann, F. Kronenberg, J. Adamski, T. Illig, *PLoS One* **2010**, *5*, e13953.
[12] O. Fiehn, W. T. Garvey, J. W. Newman, K. H. Lok, C. L. Hoppel, S. H. Adams, *PLoS One* **2010**, *5*, e15234.
[13] S. Wopereis, C. M. Rubingh, M. J. van Erk, E. R. Verheij, T. van Vliet, N. H. Cnubben, A. K. Smilde, J. van der Greef, B. van Ommen, H. F. Hendriks, *PLoS One* **2009**, *4*, e4525.
[14] P. Spégel, A. P. H. Danielsson, K. Bacos, C. L. F. Nagorny, T. Moritz, H. Mulder, K. Filipsson, *Metabolomics* **2010**, *6*, 56.
[15] N. Geidenstam, P. Spégel, H. Mulder, K. Filipsson, M. Ridderstråle, P. H. Danielsson Anders, *Obesity* **2014**, *22*, 2388.
[16] C. Morris, C. O'Grada, M. Ryan, H. M. Roche, M. J. Gibney, E. R. Gibney, L. Brennan, *PLoS One* **2013**, *8*, e72890.
[17] S. Krug, G. Kastenmuller, F. Stuckler, M. J. Rist, T. Skurk, M. Sailer, J. Raffler, W. Romisch-Margl, J. Adamski, C. Prehn, T. Frank, K. H. Engel, T. Hofmann, B. Luy, R. Zimmermann, F. Moritz, P. Schmitt-Kopplin, J. Krumsiek, W. Kremer, F. Huber, U. Oeh, F. J. Theis, W. Szymczak, H. Hauner, K. Suhre, H. Daniel, *FASEB J.* **2012**, *26*, 2607.
[18] J. E. Ho, M. G. Larson, R. S. Vasan, A. Ghorbani, S. Cheng, E. P. Rhee, J. C. Florez, C. B. Clish, R. E. Gerszten, T. J. Wang, *Diabetes* **2013**, *62*, 2689.
[19] C. I. Mack, C. H. Weinert, B. Egert, P. G. Ferrario, A. Bub, I. Hoffmann, B. Watzl, H. Daniel, S. E. Kulling, *Am. J. Clin. Nutr.* **2018**, *108*, 502.
[20] E. Pitkänen, *Clin. Chim. Acta* **1996**, *251*, 91.
[21] T. Kawasaki, N. Ogata, H. Akanuma, T. Sakai, H. Watanabe, K. Ichiyanagi, T. Yamanouchi, *Metabolism* **2004**, *53*, 583.
[22] I. Miwa, T. Taguchi, *Clin. Chim. Acta* **2013**, *422*, 42.
[23] L. Chen, C.-Y. Cheng, H. Choi, M. K. Ikram, C. Sabanayagam, G. S. Tan, D. Tian, L. Zhang, G. Venkatesan, E. S. Tai, J. J. Wang, P. Mitchell, C. M. Gemmy Cheung, R. W. Beuerman, L. Zhou, E. C. Yong Chan, T. Y. Wong, *Diabetes* **2016**, *65*, 1099.
[24] A. S. Hoefle, A. M. Bangert, A. Stamfort, K. Gedrich, M. J. Rist, Y.-M. Lee, T. Skurk, H. Daniel, *J. Nutr.* **2015**, *145*, 467.
[25] World Health Organization. Definition and diagnosis of diabetes mellitus and intermediate hyperglycaemia. Report of a WHO/IDF consultation. WHO Press 2006. Available from: https://www.who.int/publications-detail/definition-and-diagnosis-of-diabetes-mellitus-and-intermediate-hyperglycaemia.
[26] M. Willems, A. O. Quartero, M. E. Numans, *Dig. Dis. Sci.* **2001**, *46*, 2256.
[27] C. H. Weinert, B. Egert, S. E. Kulling, *J. Chromatogr. A* **2015**, *1405*, 156.
[28] M. Vinaixa, S. Samino, I. Saez, J. Duran, J. J. Guinovart, O. Yanes, *Metabolites* **2012**, *2*, 775.
[29] T. Cakir, K. R. Patil, Z. I. Onsan, K. O. Ulgen, B. Kirdar, J. Nielsen, *Mol. Syst. Biol.* **2006**, *2*, 50.
[30] R Core Team, *R: A Language and Evironment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria **2016**.
[31] T. Hothorn, F. Bretz, P. Westfall, *Biom. J.* **2008**, *50*, 346.
[32] Y. Xie, knitr: A General-Purpose Package for Dynamic Report Generation in R. R Package Version 1.15.1. https://cran.r-project.org/web/packages/knitr/index.html (**2016**).
[33] Mirai Solutions GmbH, XLConnect: Excel Connector for R. R Package Version 0.2-12. https://CRAN.R-project.org/package=XLConnect (**2016**).
[34] D. Sarkar, *Lattice: Multivariate Data Visualization with R*, Springer, New York **2008**.
[35] A. Kuznetsova, P. B. Brockhoff, R. H. B. Christensen, lmerTest: Test in Linear Mixed Effects Models. R package version 2.0-33 **2016**. https://CRAN.R-project.org/package=lmerTest
[36] H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*, Springer-Verlag, New York **2016**.
[37] B. Auguie, gridExtra: Miscellaneous Functions for "Grid" Graphics. R Package Version 2.2.1. https://CRAN.R-project.org/package=gridExtra (**2016**).
[38] C. H. Weinert, M. T. Empl, R. Krüger, L. Frommherz, B. Egert, P. Steinberg, S. E. Kulling, *Mol. Nutr. Food Res.* **2017**, *61*, 1600651.
[39] A. Bub, A. Kriebel, C. Dörr, S. Bandt, M. Rist, A. Roth, E. Hummel, S. Kulling, I. Hoffmann, B. Watzl, *JMIR Res. Protoc.* **2016**, *5*, e146.
[40] M. F. Laker, I. S. Menzies, *J. Physiol.* **1977**, *265*, 881.
[41] I. S. Menzies, *Biochem. Soc. Trans.* **1974**, *2*, 1042.
[42] T. Kellerer, B. Brandl, J. Büttner, I. Lagkouvardos, H. Hauner, T. Skurk, *Obes. Surg.* **2019**, *29*, 2132.
[43] L. H. Münger, A. Trimigno, G. Picone, C. Freiburghaus, G. Pimentel, K. J. Burton, F. P. Pralong, N. Vionnet, F. Capozzi, R. Badertscher, G. Vergères, *J. Proteome Res.* **2017**, *16*, 3321.
[44] N. Tasevska, *Nutrients* **2015**, *7*, 5816.
[45] K. M. Dungan, *Expert Rev. Mol. Diagn.* **2008**, *8*, 9.
[46] H. Sone, H. Shimano, H. Ebinuma, A. Takahashi, Y. Yano, K. T. Iida, H. Suzuki, H. Toyoshima, Y. Kawakami, Y. Okuda, Y. Noguchi, K. Ushizawa, K. Saito, N. Yamada, *Metabolism* **2003**, *52*, 1019.

[47] S. Lee, C. Zhang, M. Kilicarslan, B. D. Piening, E. Bjornson, B. M. Hallström, A. K. Groen, E. Ferrannini, M. Laakso, M. Snyder, M. Blüher, M. Uhlen, J. Nielsen, U. Smith, M. J. Serlie, J. Boren, A. Mardinoglu, *Cell Metab.* **2016**, *24*, 172.
[48] T. Kawasaki, H. Akanuma, T. Yamanouchi, *Diabetes Care* **2002**, *25*, 353.
[49] N. Friedrich, *J. Endocrinol.* **2012**, *215*, 29.
[50] W. Nowatzke, M. J. Sarno, N. C. Birch, D. F. Stickle, T. Eden, T. G. Cole, *Clin. Chim. Acta* **2004**, *350*, 201.
[51] M. J. Rist, A. Roth, L. Frommherz, C. H. Weinert, R. Krüger, B. Merz, D. Bunzel, C. Mack, B. Egert, A. Bub, B. Görling, P. Tzvetkova, B. Luy, I. Hoffmann, S. E. Kulling, B. Watzl, *PLoS One* **2017**, *12*, e0183228.
[52] K. C. Hootman, J.-P. Trezzi, L. Kraemer, L. S. Burwell, X. Dong, K. A. Guertin, C. Jaeger, P. J. Stover, K. Hiller, P. A. Cassano, *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, E4233.
[53] L. C. Eze, *Biochem. Genet.* **1989**, *27*, 487.
[54] L. C. Eze, D. A. P. Evans, *Clin. Chim. Acta* **1970**, *28*, 153.
[55] Y. L. Muller, R. L. Hanson, W. C. Knowler, J. Fleming, J. Goswami, K. Huang, M. Traurig, J. Sutherland, C. Wiedrich, K. Wiedrich, D. Mahkee, V. Ossowski, S. Kobes, C. Bogardus, L. J. Baier, *Hum. Genet.* **2013**, *132*, 697.
[56] C. Menni, E. Fauman, I. Erte, J. R. B. Perry, G. Kastenmüller, S.-Y. Shin, A.-K. Petersen, C. Hyde, M. Psatha, K. J. Ward, W. Yuan, M. Milburn, C. N. A. Palmer, T. M. Frayling, J. Trimmer, J. T. Bell, C. Gieger, R. Mohney, M. J. Brosnan, K. Suhre, N. Soranzo, T. D. Spector, *Diabetes* **2013**, *62*, 4270.
[57] F. C. Battaglia, P. J. Randle, *Biochem. J.* **1960**, *75*, 408.
[58] A. Hossain, F. Yamaguchi, T. Matsuo, I. Tsukamoto, Y. Toyoda, M. Ogawa, Y. Nagata, M. Tokuda, *Pharmacol. Ther.* **2015**, *155*, 49.
[59] T. Iida, Kishimoto Y., Y. Yoshikawa, N. Hayashi, K. Okuma, M. Tohi, K. Yagi, T. Matsuo, K. Izumori, *J. Nutr. Sci. Vitaminol.* **2008**, *54*, 511.
[60] N. Hayashi, T. Iida, T. Yamada, K. Okuma, I. Takehara, T. Yamamoto, K. Yamada, M. Tokuda, *Biosci., Biotechnol., Biochem.* **2010**, *74*, 510.
[61] N. Hayashi, T. Yamada, S. Takamine, T. Iida, K. Okuma, M. Tokuda, *J. Funct. Foods* **2014**, *11*, 152.

Supplemental Material 1

Exploring the diversity of sugar compounds in healthy, prediabetic and diabetic volunteers

Carina I. Mack, Paola G. Ferrario, Christoph H. Weinert, Björn Egert, Anja S. Hoefle, Yu-Mi Lee, Thomas Skurk, Sabine E. Kulling, Hannelore Daniel (2020):

Exploring the diversity of sugar compounds in healthy, prediabetic and diabetic volunteers

Molecular Nutrition and Food Research 2020, 64, 1901190.

https://doi.org/10.1002/mnfr.201901190

Supporting Information for online publication - File 1

1. Materials and methods

1.1. Study design and subjects

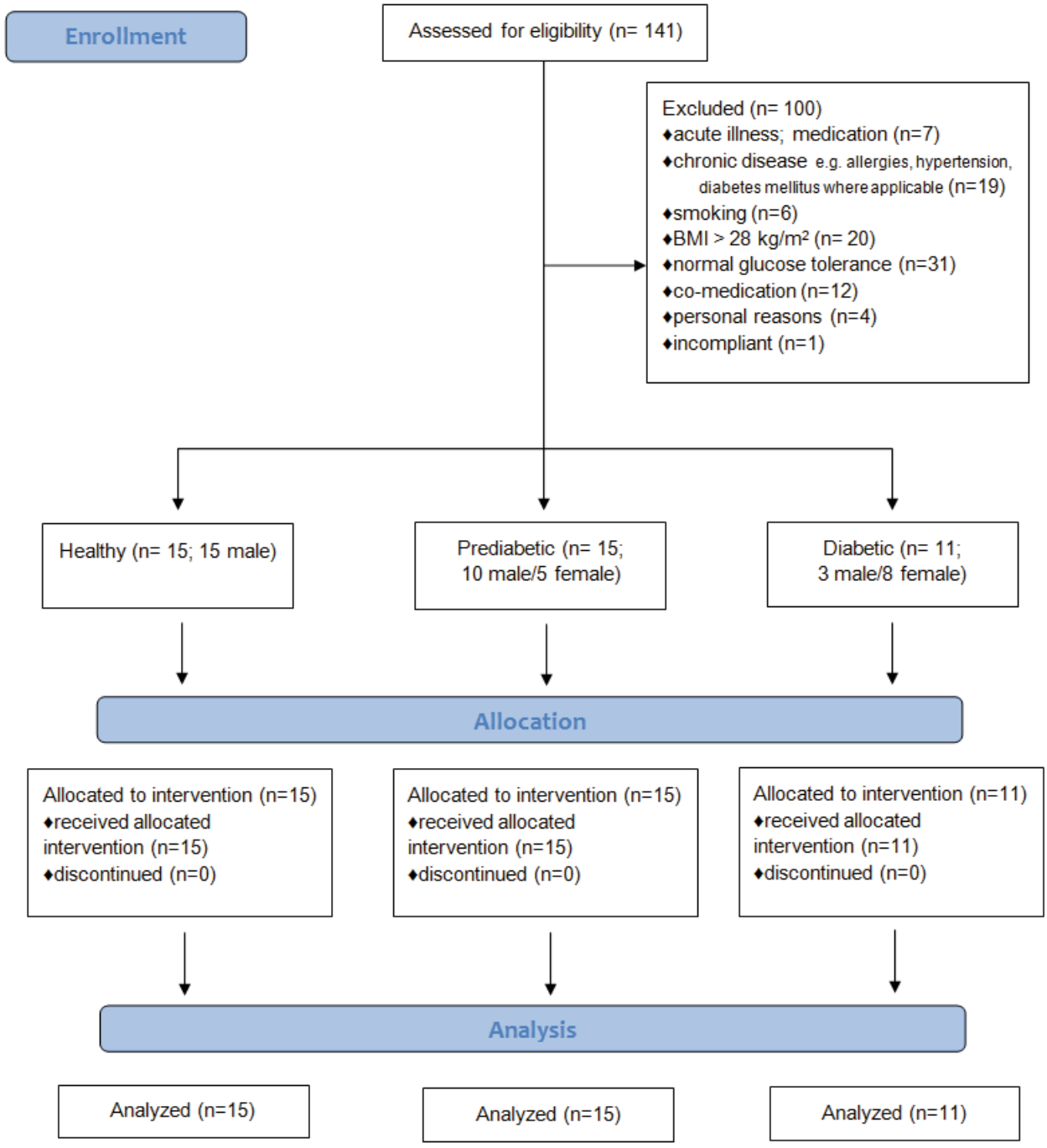


Figure S1. Flow chart including the number of individuals in the MIPROMET study screened during recruitment and reasons for non-inclusion.

Supporting Information for online publication - File 1

1.2. Semitargeted GC-MS sugar profiling

1.2.1. Plasma sample preparation

Table S1. Internal standard stock solution

Internal standard	Concentration (µmol/L in H_2O)	Manufacturer	Purity
1-*O*-Methyl-2-deoxy-D-ribose	175	Sigma-Aldrich	96%
Phenyl-*β*-D-glucopyranoside	85	Sigma-Aldrich	95%
D-Pinitol	175	ABCR	99%

1.2.2. GC-MS analysis

Table S2. GC-MS instrument and components

Instrument component / software	Name	Manufacturer
Gas chromatograph	GC-2010	Shimadzu Corp, Kyoto, Japan
Mass spectrometer	QP2010 Ultra	Shimadzu Corp, Kyoto, Japan
Auto sampler	AOC-20s	Shimadzu Corp, Kyoto, Japan
PTV Injector	OPTIC-4	GL Sciences, Eindhoven, The Netherlands
GCMS instrument software	GCMSSolution 4.11	Shimadzu Corp, Kyoto, Japan
PTV software	Evolution Workstation 4.1	GL Sciences, Eindhoven, The Netherlands

Table S3. Method parameters; whenever measurement parameters differed during measurement of KarMeN fasting plasma samples, the changed parameters are listed separately.

Parameter	Setting / value
GC parameters	
Carrier gas	Helium
GC mode	MIPROMET: Constant pressure (256 kPa)
	KarMeN: Constant pressure (180 kPa)
Liner type	Deactivated split liner with quartz wool
Column	MIPROMET: Restek Rxi-5SilMS; L = 60 m plus 10 m of an integrated pre-column; d_c = 0.25 mm; d_f = 0.25 µm
	KarMeN: Restek Rxi-5SilMS; L = 57 m plus 2 m of an integrated pre-column; d_c = 0.25 mm; d_f = 0.25 µm
GC temperature ramp	150 °C → 4 °C/min → 180 °C → 7 °C/min → 200 °C → 2.25 °C/min → 235 °C → 10 °C/min → 270 °C → 5.25 °C/min → 310 °C → 20 °C/min → 330 °C (3,97 min). Run time: 42 min.
Injection mode	Cold split
Injection volume	1.2 µL
Split ratio	1:5
PTV temperature ramp	90 °C → 60 °C/s→ 280 °C, hold until end of run
Interface temperature	300 °C

Supporting Information for online publication - File 1

Table S3 continued.

	MS parameters
Ion source temperature	200 °C
Ionization mode	EI (70 eV)
MS Mode	Scan/SIM
Scan speed	20.000 u/s
Scan range	m/z 60-600
Event time (scan)	40 ms
SIM time frame 1 (6.50-15.10 min)	MIPROMET: Event time: 620 ms; m/z 147.05, 204.10, 217.10, 307.15, 319.20, 117.00, 160.10, 292.15, 277.20, 333.10, 245.00, 218.00, 329.00, 314.00, 220.00, 310.00, 104.00, 323.00, and 295.00
(6.50-15.86 min)	KarMeN: Event time: 620 ms; m/z 204.10, 217.10, 307.15, 319.20, 117.00, 160.10, 292.15, 277.20, 333.10, 245.00, 218.00, 329.00, 220.00, and 295.00
SIM time frame 2 (15.10 - 20.17 min)	MIPROMET: Event time: 680 ms; m/z 147.05, 204.10, 217.10, 307.15, 319.20, 117.00, 160.10, 318.20, 205.10, 333.10, 361.15, 220.00, 202.10, 260.00, 345.00, 255.00, 156.00, 206.00, 105.00, 310.15, 323.70, 162.05, 104.15, 337.00, and 308.00
(15.86-21.19 min)	KarMeN: Event time: 680 ms; m/z 204.10, 217.10, 307.15, 319.20, 117.00, 160.10, 318.20, 205.10, 333.10, 361.15, 220.00, 202.10, 260.00, 345.00, 308.00, and 156.00
SIM time frame 3 (20.17-31.75 min)	MIPROMET: Event time: 600 ms; m/z 147.05, 204.10, 217.10, 307.15, 319.20, 318.20, 205.10, 333.10, 299.00, 202.10, 361.15, 441.00, 382.00, 357.00, 321.30, 220.20, 192.15, and 266.20
(21.19-32.57 min)	KarMeN: Event time: 620 ms; m/z 204.10, 217.10, 307.15, 319.20, 318.20, 205.10, 333.10, 299.00, 202.10, 361.15, 441.00, 382.00, 220.20, and 192.15
SIM time frame 4 (31.75-41.99 min)	MIPROMET: Event time: 400 ms; m/z 147.05, 204.10, 217.10, 307.15, 319.20, 361.15, 367.00, 442.00, 220.00, 174.00
(32.57-41.99 min)	KarMeN: Event time: 400 ms; m/z 204.10, 217.10, 307.15, 319.20, and 361.15

Supporting Information for online publication - File 1

1.2.3. Data treatment and evaluation

Table S4. Integration and identification parameters

Parameter	Setting/value
Slope	100/min
Width	1 s
Drift	0/min
T.DBL	1000 min
Minimal area	0
Smoothing	none
Default bandtime	0.1 min
Spectrum confirmation	use ratios of target and reference ions[a]
Reference ion mode	absolute
Ratio deviation allowance	30%*

[a]In some cases it became necessary to use a higher allowance, or to use only one or no reference ion due to overlapping peaks with similar MS spectra.

After automatic integration a manual check of peak assignment and occurrence of coelution was performed. The repeatability was checked by CV's of QC samples and the internal standards.

1.3. Study design and subjects

Table S5. Participant characteristics (range and mean in brackets) in the KarMeN study.

	n	Sex (n)		Age (years)			BMI (kg/m²)		
		Male	Female	m	f	all	m	f	all
Male	28	-	-	-	-	21-80 (50.6)	-	-	20.0-29.9 (24.9)
Female	30	-	-	-	-	23-80 (50.4)	-	-	19.4-29.2 (23.6)
All	58	28	30	-	-	21-80 (50.5)	-	-	19.4-29.9 (24.2)
Age (21-30 years)	10	5	5	21-30 (25.4)	23-29 (26.0)	21-30 (25.7)	20.0-29.8 (24.3)	20.1-22.3 (21.1)	20.0-29.8 (22.7)
Age (50-72 years)	23	11	12	50-71 (61.2)	50-72 (61.7)	50.72 (61.5)	22.5-29.2 (25.2)	20.9-29.2 (24.9)	20.9-29.2 (25.0)
BMI (19-23.9 kg/m²)	29	12	17	23-71 (44.8)	23-69 (40.7)	23-71 (42.4)	20.0-23.9 (22.5)	19.4-23.8 (21.5)	19.4-23.9 (21.9)
BMI (24-30 kg/m²)	29	16	13	21-80 (54.9)	42-80 (63.0)	21-80 (58.5)	24.0-29.9 (26.7)	24.0-29.2 (26.3)	24.0-29.9 (26.5)

1.4. Analysis of the ingested test drink for the OGTT and further commercial OGTT solutions

In addition to the GC instrument and components listed in Supporting Information Table S3, the ZX2 modulator from ZOEX Corp. (Houston, USA) was used for the GC×GC-MS measurement. The method parameters are listed in Supporting Information Table S6. The GC×GC-MS chromatograms were visualized using GC Image™ GC×GC Software version 2.5 (GCImage, LLC Lincoln NE, USA).

Supporting Information for online publication - File 1

Table S6. Method parameters for GC×GC-MS measurement.

Parameter	Setting / value
GC parameters	
Carrier gas	Helium
GC mode	Constant velocity (29.7 cm/s)
Liner type	Deactivated split liner with quartz wool
^{1}D-Column	Restek Rxi-5SilMS; ^{1}L = 30 m plus 10 m of an integrated pre-column; $^{1}d_c$ = 0.25 mm; $^{1}d_f$ = 0.25 µm
^{2}D-Column	SGE BPX50; $^{2}L_{total}$ = 2.2 m including a "separation segment" of $^{2}L_{sep}$ = 0.7 m; $^{2}d_c$ = 0.15 mm; $^{2}d_f$ = 0.15 µm
Column connector	SilTite MiniUnion (Trajan/SGE)
GC temperature ramp	90 °C → 3 °C/min → 300 °C → 15 °C/min → 320 °C (7.67 min). Run time: 80 min.
Injection mode	Cold split
Injection volume	1 µL
Split ratio	1:10
PTV temperature ramp	90 °C → 60 °C/s→ 280 °C, hold until end of run
Interface temperature	300 °C
Modulation parameters	
Modulator type	Cryogenic, air-based, loop-type
Modulation period (P_M)	3 s
Cold jet temperature	-90 °C
Hot jet temperature	Programmed stepwise, at least 100 °C above oven temperature
Hot jet duration	250 ms
MS parameters	
Ion source temperature	200 °C
Ionization mode	EI (70 eV)
MS Mode	Scan
Scan speed	20.000 u/s
Scan range	m/z 56-550
Event time	30 ms

Supporting Information for online publication - File 1

2. Results

Table S7. Summary of all measured known and unknown (U) sugar compounds, occurrence of partial or total chromatographic coelution, the corresponding target and reference ions, their ratios, the median, minimum and maximum peak area in study samples, frequency in all study samples, and intra- and inter-day repeatabilities (before and after signal intensity drift/batch correction) in QC samples and intermediate precision of internal standards over all samples (QC plus study samples).

Sugar compounds[a]	Coelution[b]	Target ion	Reference ions	Ratios target/reference ions[c]	Median peak area	Minimum peak area	Maximum peak area	Detection frequency (%)	Intra-day$_{uncorr}$ (mean)[d]	Inter-day$_{uncorr}$	Intra-day$_{corr}$ (mean)[d]	Inter-day$_{corr}$
Xylose	partial$_{chrom}$	307.15	204.10, 117.00	551.59, 2477.54	1.38E+04	6.44E+03	5.38E+04	100.0	14.7	24.5	6.0	5.5
Arabinose		307.15	117.00, 204.10	4096.49, 577.05	2.28E+04	1.35E+04	4.41E+04	100.0	13.4	24.5	6.6	6.0
Psicose	partial$_{mass}$	307.15	117.00, 205.10	107.59, 25.49	2.84E+04	3.37E+03	1.99E+05	100.0	16.1	24.8	7.6	6.8
Fructose[e]		307.15	205.10, 117.00	19.25, 23.67	9.25E+04	1.84E+04	5.34E+05	100.0	18.9	38.5	6.8	6.4
Mannose		319.20	205.10, 160.10	57.41, 35.51	3.07E+06	1.03E+06	6.80E+06	100.0	14.4	24.5	6.7	5.7
Glucose[e]		319.20	205.10, 160.10	49.79, 31.68	3.50E+07	1.96E+07	6.75E+07	100.0	7.3	13.6	3.4	2.8
Sedoheptulose[e]		319.20	205.10	110.76	9.81E+03	5.16E+03	3.87E+04	100.0[f]	16.1	25.7	7.0	6.3
Sucrose		361.15	319.20, 307.15	5.21, 0.51	9.42E+04	2.59E+04	1.76E+06	100.0	34.8	58.0	6.3	6.0
Lactose[e]		361.15	307.15, 319.20	23.03, 46.14	1.46E+04	3.84E+03	4.03E+05	100.0[f]	18.8	27.8	7.0	7.2
Lactulose[e]		361.15	204.10, 307.15	229.84, 13.98	3.02E+04	5.60E+03	6.89E+05	66.1	17.8	33.3	10.5	8.8
Maltose[e]		361.15	319.20, 204.10	20.42, 57.63	6.12E+04	9.00E+03	2.32E+06	100.0	43.2	55.1	21.2	17.8
Trehalose	partial$_{chrom}$	361.15	319.20, 307.15	0.29, 0.11	2.36E+04	8.93E+03	9.98E+04	100.0	23.2	39.3	12.2	11.0
Threitol		204.10	160.10, 117.00	145.47, 328.36	9.66E+03	4.13E+03	4.90E+04	100.0[f]	12.2	16.7	7.6	6.0
meso-Erythritol		204.10	117.00, 307.15	237.66, 59.72	3.17E+04	1.61E+04	8.67E+04	100.0	13.4	21.1	7.5	6.5
Arabitol		307.15	319.20, 117.00	104.80, 304.61	3.37E+04	1.55E+04	9.47E+04	100.0	13.2	23.0	6.9	5.8
Ribitol	partial$_{chrom}$	307.15	319.20, 204.10	11.00, 47.72	4.93E+03	2.61E+03	1.08E+04	100.0[f]	12.9	20.2	3.6	3.3
1,5-Anhydroglucitol	partial$_{chrom}$	260.00	319.20, 117.00	138.81, 326.62	1.64E+05	3.57E+04	4.52E+05	100.0	13.4	21.8	7.7	6.0
Mannitol	partial$_{chrom}$	319.20	205.10, 117.00	47.70, 13.30	1.29E+05	2.71E+04	9.90E+05	100.0	19.6	25.1	10.7	10.7
scyllo-Inositol	partial$_{chrom}$	318.20	307.15, 204.10	25.76, 1307.79	9.33E+04	3.61E+02	3.31E+05	100.0	14.0	25.1	7.0	5.6
myo-Inositol	partial$_{chrom}$	318.20	204.10, 192.15	79.61, 21.14	8.03E+05	5.46E+05	1.70E+06	100.0	14.1	24.1	7.1	5.5
Sorbitol	partial$_{chrom}$	319.20	307.15	10.50	2.67E+04	1.07E+04	2.63E+05	100.0	17.1	30.0	8.1	7.5
Erythronic acid		292.15	117.00, 319.20	74.53, 9.81	8.71E+04	1.33E+04	1.65E+05	100.0	14.2	23.8	7.8	6.1
Threonic acid	partial$_{chrom}$	292.15	117.00, 319.20	35.78, 12.41	3.79E+05	6.72E+04	9.42E+05	100.0	18.0	30.3	8.3	6.9
Glucuronic acid[e]	partial$_{chrom}$	333.10	205.10, 160.10	75.48, 82.80	6.06E+04	3.13E+04	2.36E+05	100.0	14.4	24.5	6.8	6.2
Xylonic acid		292.15	333.10	22.92	1.92E+03	7.23E+02	1.55E+04	100.0[f]	27.7	36.7	15.0	12.7
Ribonic acid		292.15	333.10	42.52	2.32E+03	8.87E+02	7.03E+03	100.0[f]	31.7	36.0	17.7	16.4

Supporting Information for online publication - File 1

Arabonic acid		292.15	333.10	25.24	6.41E+03	8.96E+02	4.24E+04	100.0[f]	36.7	43.7	19.5	15.8
U16		117.00	204.10, 160.10	18.46, 3.83	1.95E+04	1.73E+03	2.11E+05	99.2	52.2	93.5	15.7	15.0
U01		277.20	292.15, 117.00	43.45, 2.59	1.01E+05	3.06E+04	2.24E+05	100.0	37.2	60.9	12.5	10.4
U03	$partial_{mass}$	361.15	319.20, 204.10	10.20, 39.56	9.57E+04	2.85E+04	3.13E+05	100.0	16.8	24.7	9.1	7.2
U07		319.20	117.00, 204.10	187.95, 106.63	9.34E+03	3.59E+03	3.29E+04	100.0[f]	26.1	32.5	15.1	13.4
U08		307.15	319.20, 333.10	146.17, 225.20	3.01E+04	1.06E+04	1.12E+05	100.0	31.5	37.7	14.0	12.7
U10		361.15	299.00, 204.10	0.89, 1.83	3.34E+04	5.32E+03	1.59E+06	100.0	65.4	78.2	33.8	30.8
U11		361.15	204.10, 299.00	25.22, 7.59	8.26E+03	2.72E+02	1.23E+06	100.0[f]	101.3	108.6	41.8	38.4
U12		202.10	333.10, 319.20	16.02, 14.37	5.24E+04	8.49E+03	2.95E+05	100.0	15.4	24.3	9.2	6.8
U15	$partial_{chrom}$	319.20	205.10, 217.10	117.18, 217.79	6.02E+03	1.91E+03	2.06E+04	100.0[f]	22.3	36.2	14.8	13.9
U17		292.15	333.10	27.96	4.19E+03	1.42E+03	1.48E+04	100.0[f]	24.4	32.9	12.2	10.3
U18		117.00	204.10, 217.10	46.34, 18.22	7.71E+04	3.94E+03	4.44E+05	90.9	13.2	20.4	8.5	7.0
U19		319.20	204.10	23.42	7.29E+03	3.06E+02	3.56E+05	100.0[f]	90.6	93.3	47.9	48.7
U20		319.20	204.10	187.22	2.28E+03	1.93E+01	4.70E+05	98.3[f]	117.8	116.9	47.4	47.2
Methyl-2-deoxyribose [e, g]	$partial_{chrom}$	160.10	117.00, 204.10	4383.23, 197.83				100.0	9.7	16.5	8.7	8.9
Pinitol[g]		260.00	318.20, 204.10	77.15, 16.16				100.0	9.0	20.0	7.8	8.2
Phenyl-β-glucopyranoside[g]		361.15	319.20, 204.10	7.31, 7.71				100.0	10.5	22.5	8.4	9.0

[a]All named sugar compounds were unambiguously identified using co-chromatography with standard substances; the remaining unknowns could be assigned to a substance class according to mass spectra. [b]$partial_{mass.}$: total chromatographic coelution, separation through masses possible; $partial_{chrom.}$: partial chromatographic coelution and separation through masses possible. [c]Ratios of reference ions to target ion, deviation allowance was usually 30%. [d]Mean of 5 measurement days intra-day precision. [e]Only first of two derivatives is shown. [f]Trace analytes. [g]Internal standards.

Supporting Information for online publication - File 1

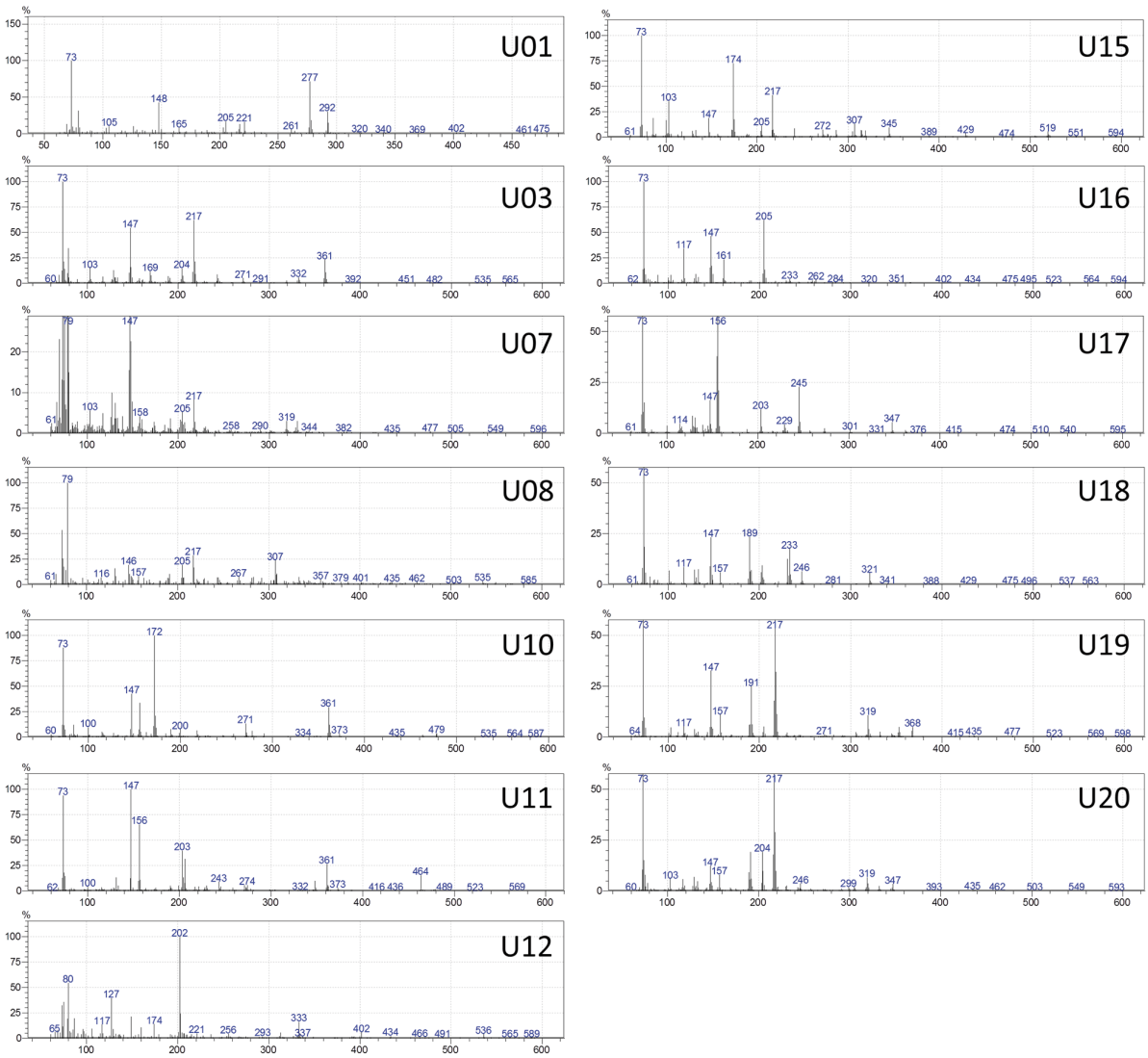


Figure S2. Mass spectra of the unknown sugar-like compounds.

Supporting Information for online publication - File 1

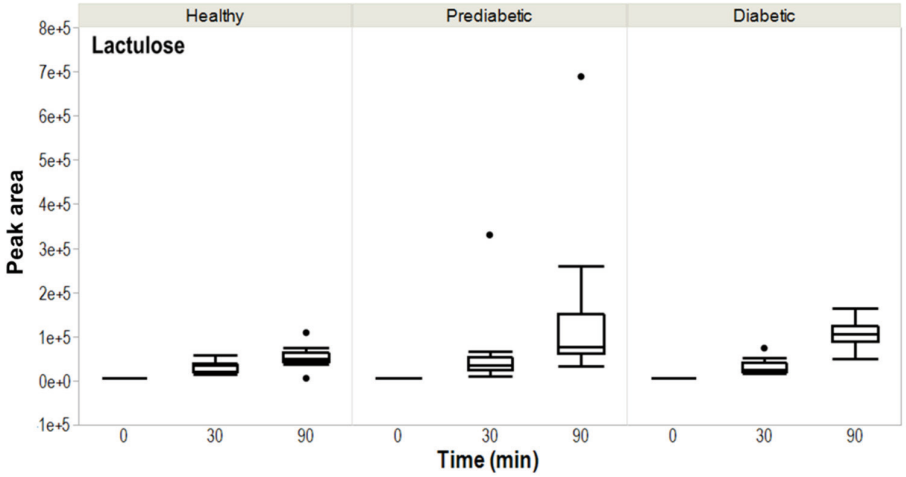


Figure S3. Boxplots for the measured time points (0, 30 and 90 min) of lactulose grouped into healthy (n = 15), prediabetic (n = 15) and diabetic (n = 11) volunteers.

Supporting Information for online publication - File 1

2.1. Verification of findings in fasting plasma samples in an independent study

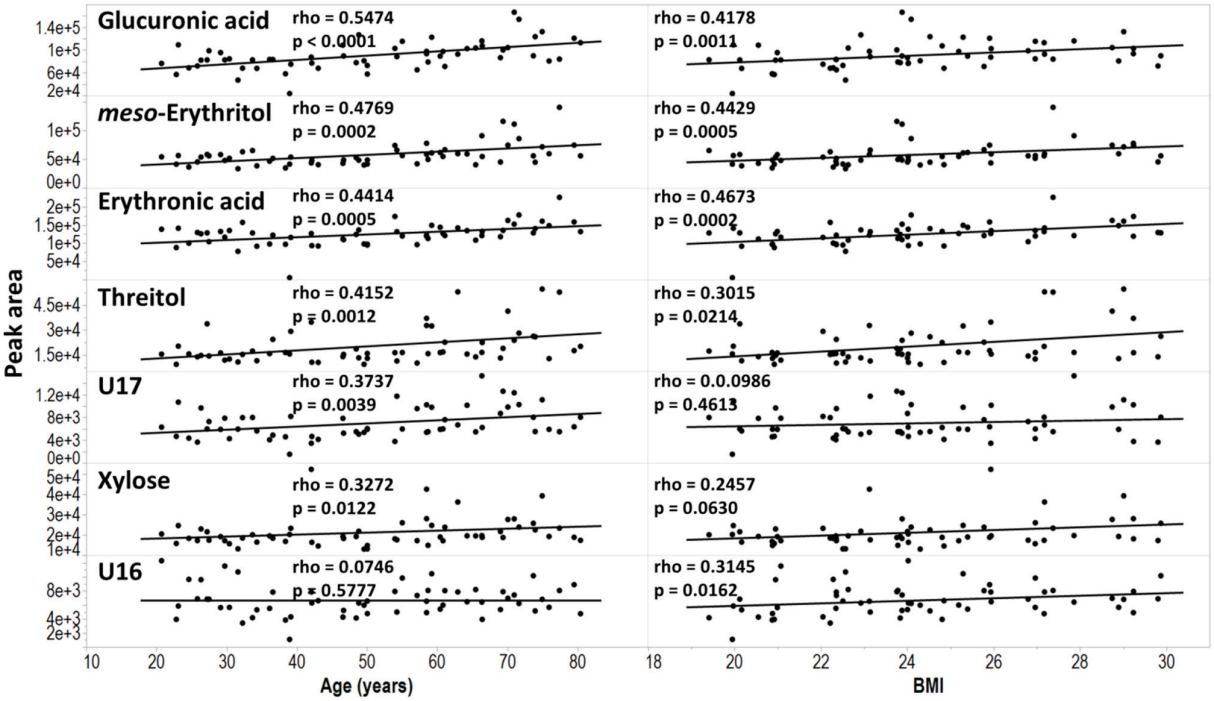


Figure S4. Scatterplots of glucuronic acid, *meso*-erythritol, threitol, and xylose showing significant Spearman correlation coefficients with age and/or BMI in KarMeN fasting plasma samples (n = 58).

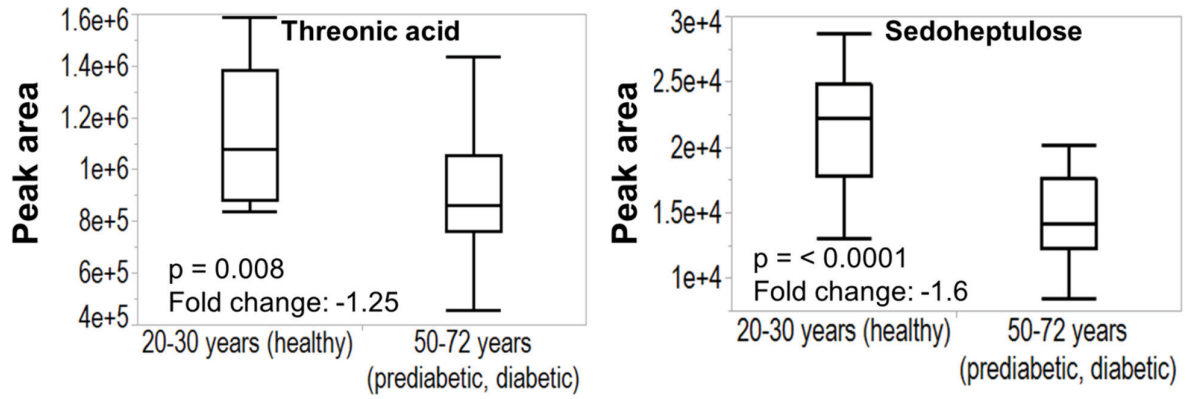


Figure S5. Boxplots of threonic acid and sedoheptulose with significant differences between 20-30 years old (n = 10) and 50-72 years old (n = 23) KarMeN participants.

Supporting Information for online publication - File 1

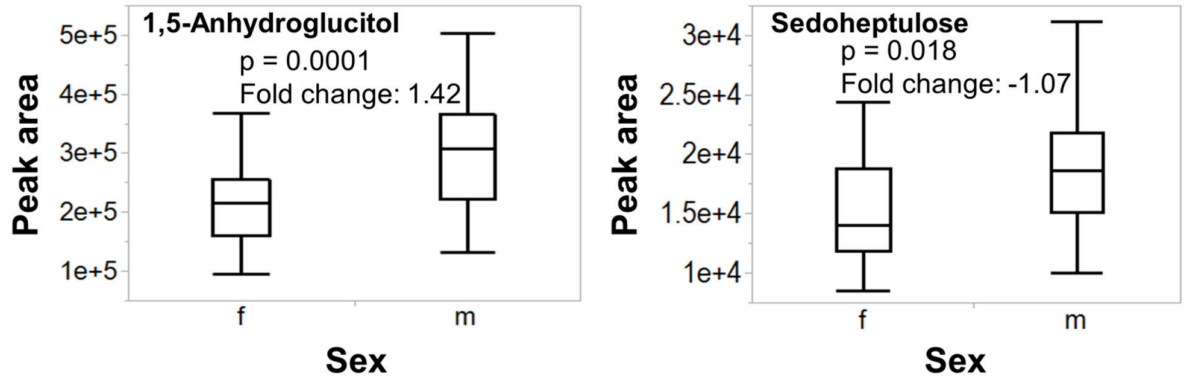


Figure S6. Boxplots of sedoheptulose and 1,5-anhydroglucitol with significant differences between male (n = 28) and female (n = 30) KarMeN participants.

Supplemental Material 2

Exploring the diversity of sugar compounds in healthy, prediabetic and diabetic volunteers

Carina I. Mack, Paola G. Ferrario, Christoph H. Weinert, Björn Egert, Anja S. Hoefle, Yu-Mi Lee, Thomas Skurk, Sabine E. Kulling, Hannelore Daniel (2020):

Exploring the diversity of sugar compounds in healthy, prediabetic and diabetic volunteers

Molecular Nutrition and Food Research 2020, 64, 1901190.

https://doi.org/10.1002/mnfr.201901190

Online Supporting Material File 2

The plasma "sweet" component: Exploring the diversity of sugar compounds in healthy, prediabetic and diabetic volunteers

Statistical Analyses

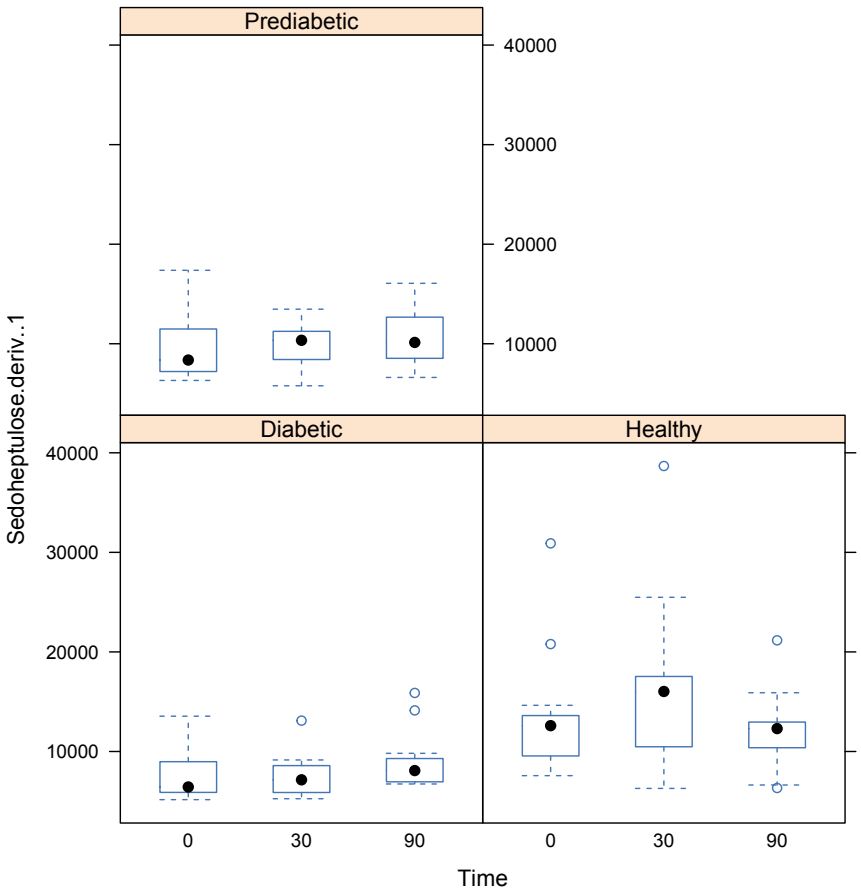
Aim

The aim is to analyse the longitudinal data of the metabolites in dependence of the variable "group" and the variable "time", after finding a proper fitting model. Interactions terms betwenn group and time are considered, as well as models corrected for the confounders "age" and "BMI". A general model is proposed and compared with the simpler model considering only "group" and "time" as independent variables.

```
library("knitr")
library("XLConnect")
library("lattice")
library("lmerTest")
library("multcomp")
```

1 Sedoheptulose.deriv..1: Example of potential confounding by age

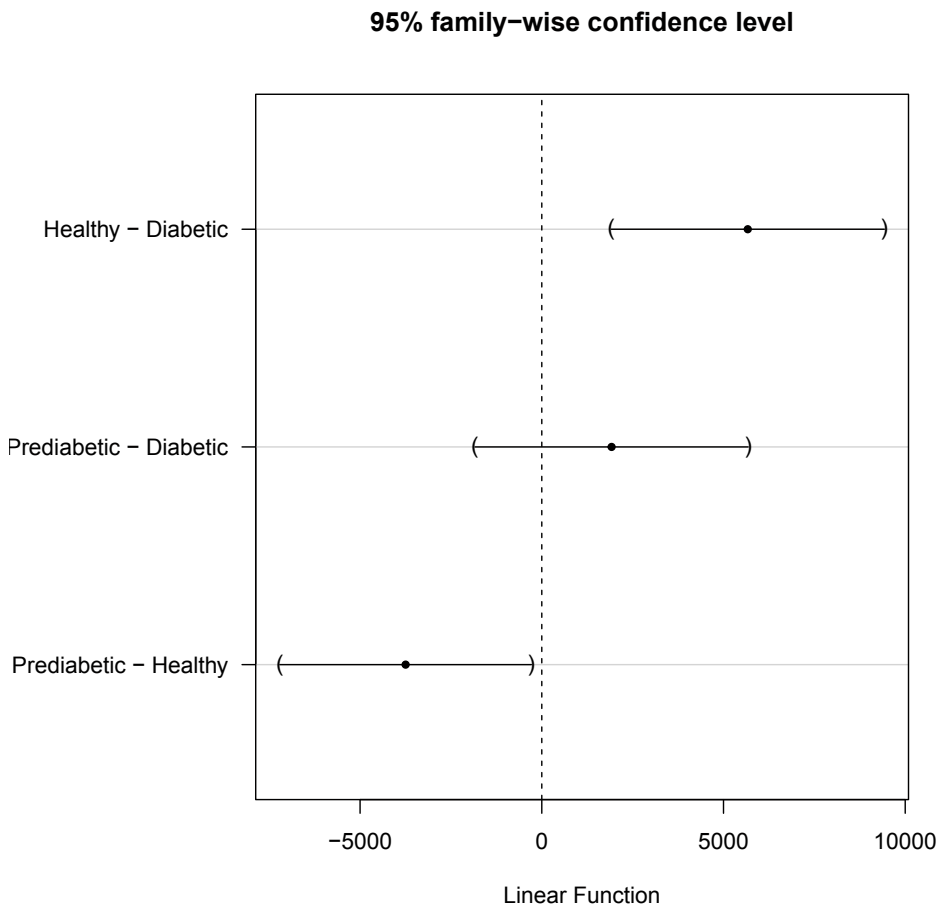
```
print(bwplot(Sedoheptulose.deriv..1~as.factor(Time..min.)|Group,data,xlab="Time"))
```



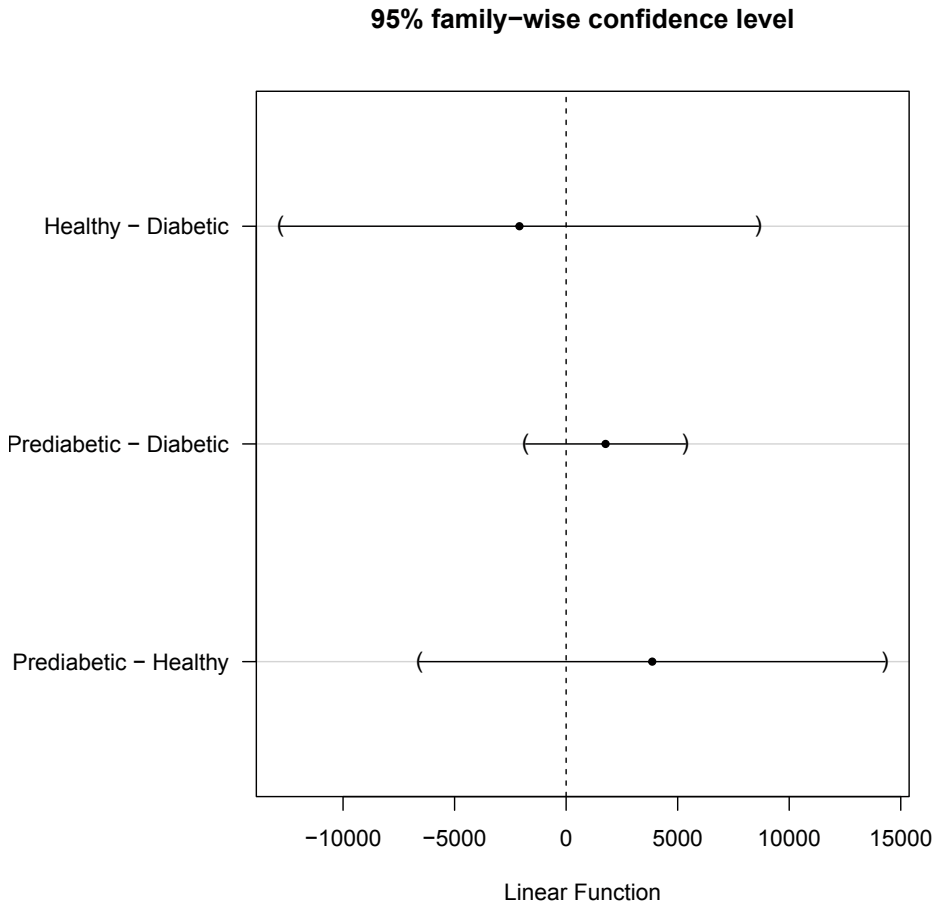

```
mod1<-lmer(Sedoheptulose.deriv..1~Group+Time..min.+(1|Participants),data)
```

2

```
mod2<-lmer(Sedoheptulose.deriv..1~Group+Time..min.+(1|Participants)+Age..years.,data)
anova(mod1,mod2)

## refitting model(s) with ML (instead of REML)

## Data: data
## Models:
## object: Sedoheptulose.deriv..1 ~ Group + Time..min. + (1 | Participants)
## ..1: Sedoheptulose.deriv..1 ~ Group + Time..min. + (1 | Participants) +
## ..1:     Age..years.
##        Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## object  7 2312.3 2331.8 -1149.1   2298.3
## ..1     8 2311.0 2333.3 -1147.5   2295.0 3.2941      1    0.06953 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

res1<-glht(mod1,linfct=mcp(Group="Tukey"))
par("mai" = par("mai") * c(1, 2.2, 1, 1))
plot(res1)
```



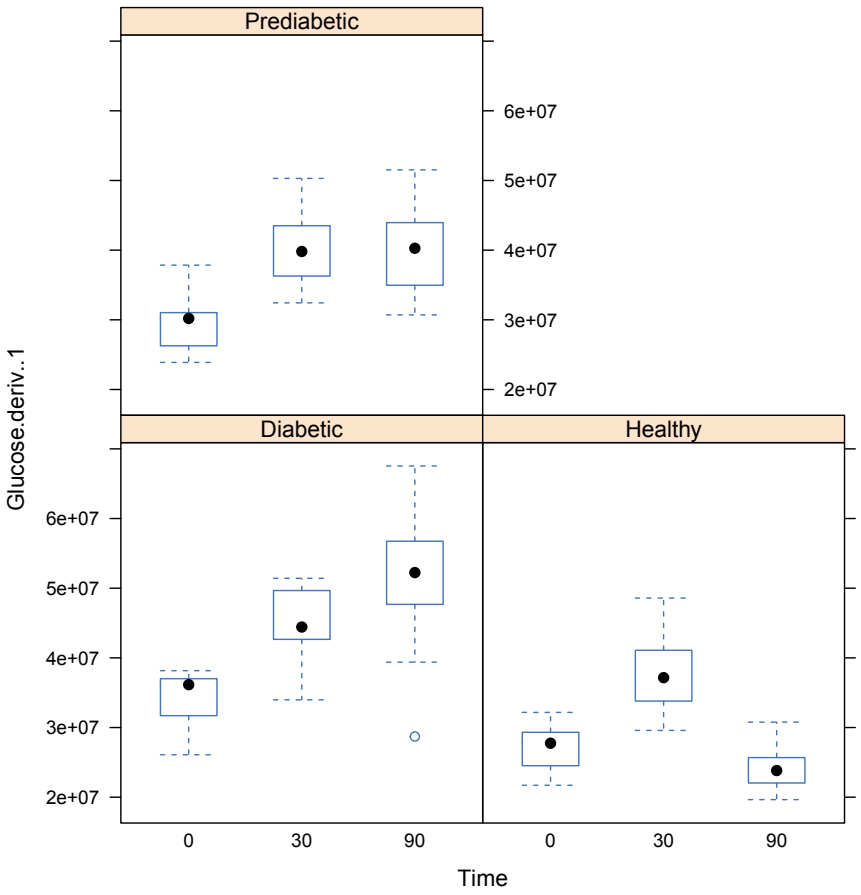

```
res2<-glht(mod2,linfct=mcp(Group="Tukey"))
plot(res2)
```


95% family−wise confidence level
Healthy − Diabetic
Prediabetic − Diabetic
Prediabetic − Healthy
−10000
−5000
0
5000
10000
15000
Linear Function

5

2 Glucose.deriv..1: Example of model considering interaction terms

```
print(bwplot(Glucose.deriv..1~as.factor(Time..min.)|Group,data,xlab="Time"))
```




```
mod1<-lmer(Glucose.deriv..1~Group*Time..min.+(1|Participants),data)
```

```
mod2<-lmer(Glucose.deriv..1~Group+Time..min.+(1|Participants),data)
anova(mod1,mod2)

## refitting model(s) with ML (instead of REML)

## Data: data
## Models:
## ..1: Glucose.deriv..1 ~ Group + Time..min. + (1 | Participants)
## object: Glucose.deriv..1 ~ Group * Time..min. + (1 | Participants)
##        Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## ..1     7 4152.4 4171.9 -2069.2   4138.4
## object 11 4083.4 4114.1 -2030.7   4061.4 77.027      4  7.423e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The choice of the proper test remains open.

3 Threitol: Example of general model considering interaction terms and potential confounders

```
print(bwplot(Threitol~as.factor(Time..min.)|Group,data,xlab="Time"))
## subject-specific slopes, intercepts as random effects; interactions
mod1<-lmer(Threitol~Group*Time..min.+(Time..min.|Participants)
           +Age..years.+BMI..kg.mš.,data)
## subject-specific slopes and intercepts as random effects
mod2<-lmer(Threitol~Group+Time..min.+(Time..min.|Participants)
           +Age..years.+BMI..kg.mš.,data)
##  only intercepts as random effects, interactions
mod3<-lmer(Threitol~Group*Time..min.+(1|Participants)
           +Age..years.+BMI..kg.mš.,data)
##  only intercepts as random effects
mod4<-lmer(Threitol~Group+Time..min.+(1|Participants)
           +Age..years.+BMI..kg.mš.,data)
anova(mod1,mod2)
anova(mod1,mod3)
anova(mod2,mod4)
anova(mod3,mod4)
```

Such a model could be considered in a future, larger study.

Supplemental Material 3

Exploring the diversity of sugar compounds in healthy, prediabetic and diabetic volunteers

Carina I. Mack, Paola G. Ferrario, Christoph H. Weinert, Björn Egert, Anja S. Hoefle, Yu-Mi Lee, Thomas Skurk, Sabine E. Kulling, Hannelore Daniel (2020):

Exploring the diversity of sugar compounds in healthy, prediabetic and diabetic volunteers

Molecular Nutrition and Food Research 2020, 64, 1901190.

https://doi.org/10.1002/mnfr.201901190

First published online 13th March, 2020 by Wiley-VCH Verlag GmbH & Co. KGaA. Reproduced with kind permission by the publisher.

The following seven pages are an excerpt of the Online Supporting Material 3. This excerpt describes the statistical reports and the R scripts. The compilation is the same for all other sugar compounds. The complete Online Supporting Material 3 can be accessed at: https://onlinelibrary.wiley.com/doi/10.1002/mnfr.201901190. All essential information of this file is summarized in Table 2 and Figure 3 of the main manuscript (see above).

Online Supporting Material File 3 of

The plasma "sweet" component: Exploring the diversity of sugar compounds in healthy, prediabetic and diabetic volunteers

Statistical Analyses

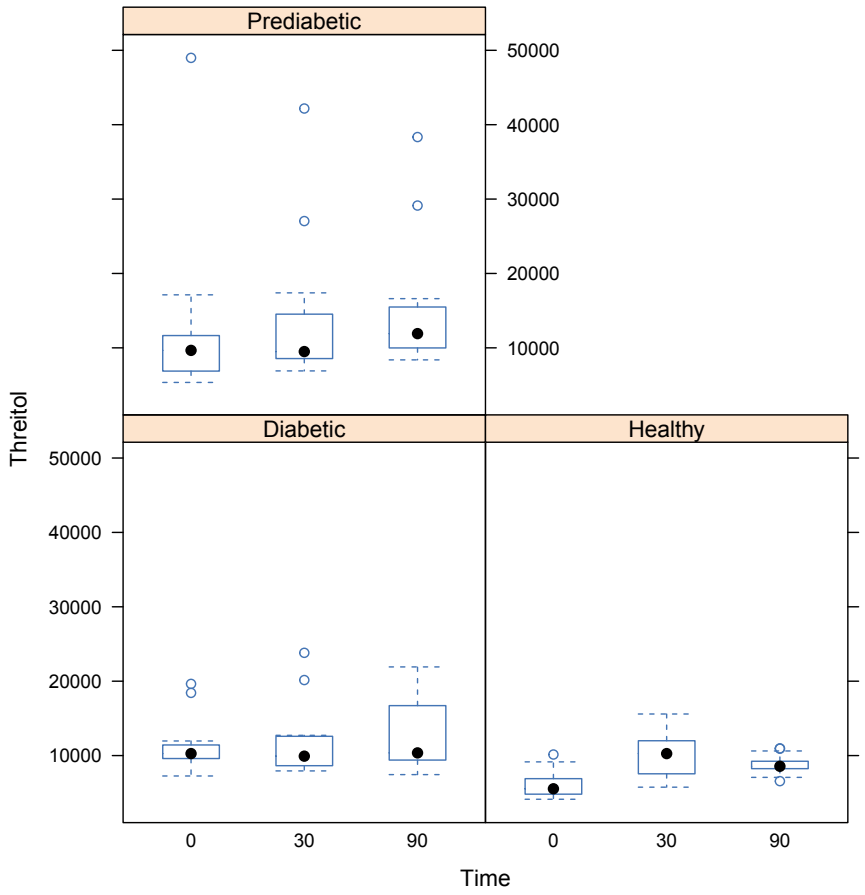
Aim

The aim is to screen sugar compounds associated with health status (healthy, prediabetic, diabetic)

```
library("knitr")
library("XLConnect")
library("lattice")
library("lmerTest")
library("multcomp")
```

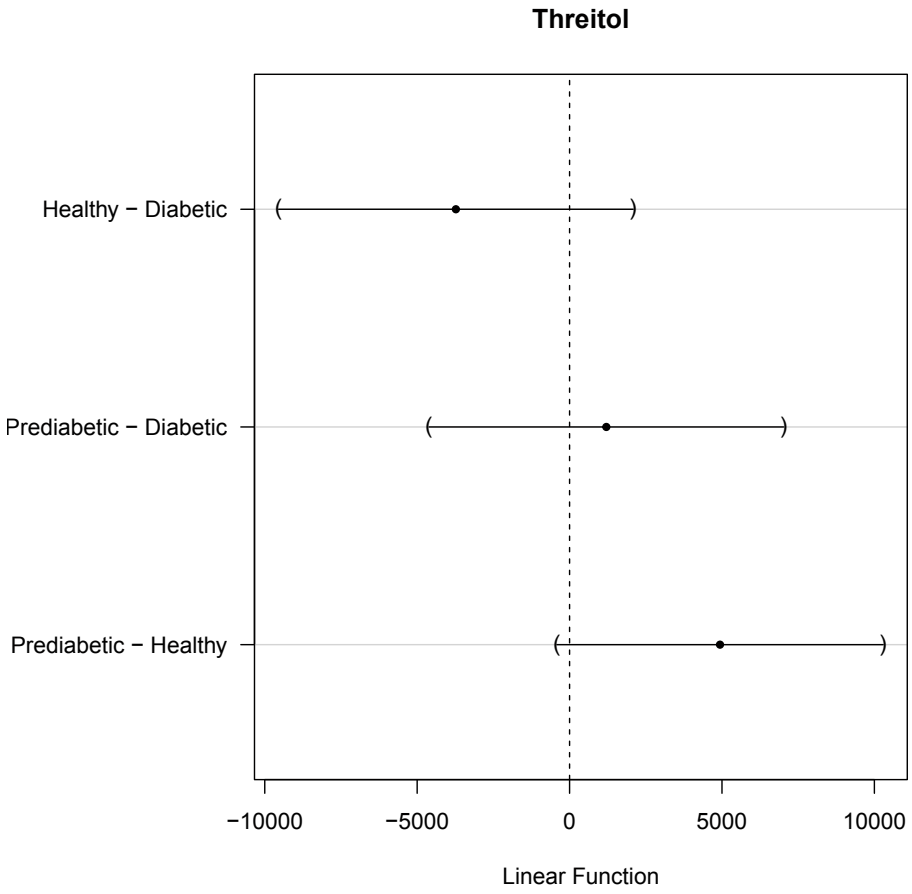
1 Threitol

```
print(bwplot(Threitol~as.factor(Time..min.)|Group,data,xlab="Time"))
```




```
mod1<-lmer(Threitol~Group+Time..min.+(1|Participants),data)
res<-glht(mod1,linfct=mcp(Group="Tukey"))
```

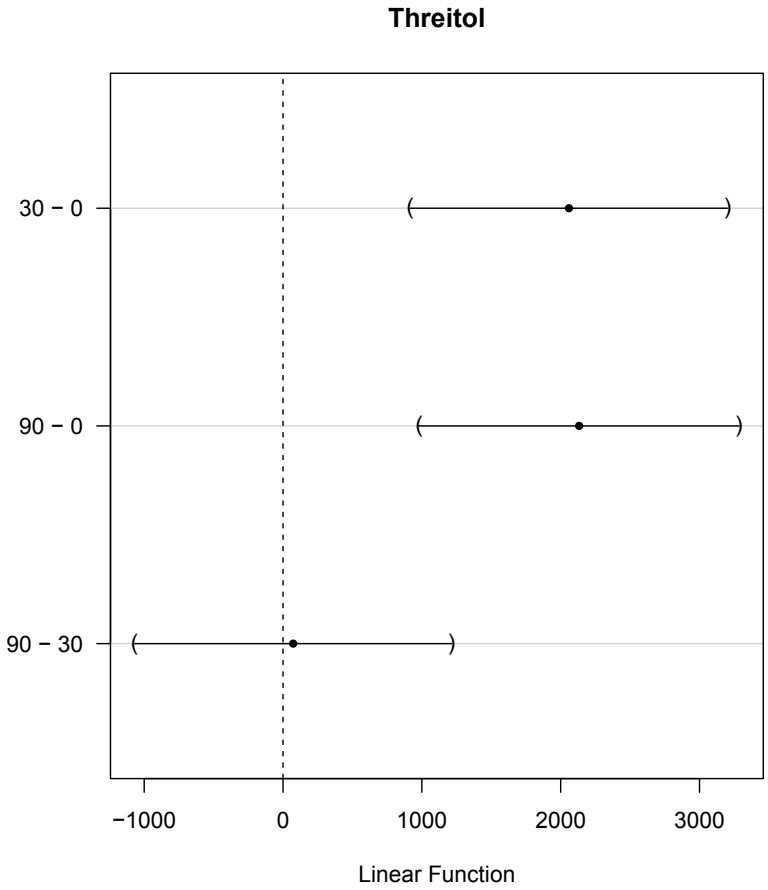
```
par("mai" = par("mai") * c(1, 2.2, 1, 1))
plot(res,main="Threitol")
```




```
summary(res)

##
##   Simultaneous Tests for General Linear Hypotheses
```

3

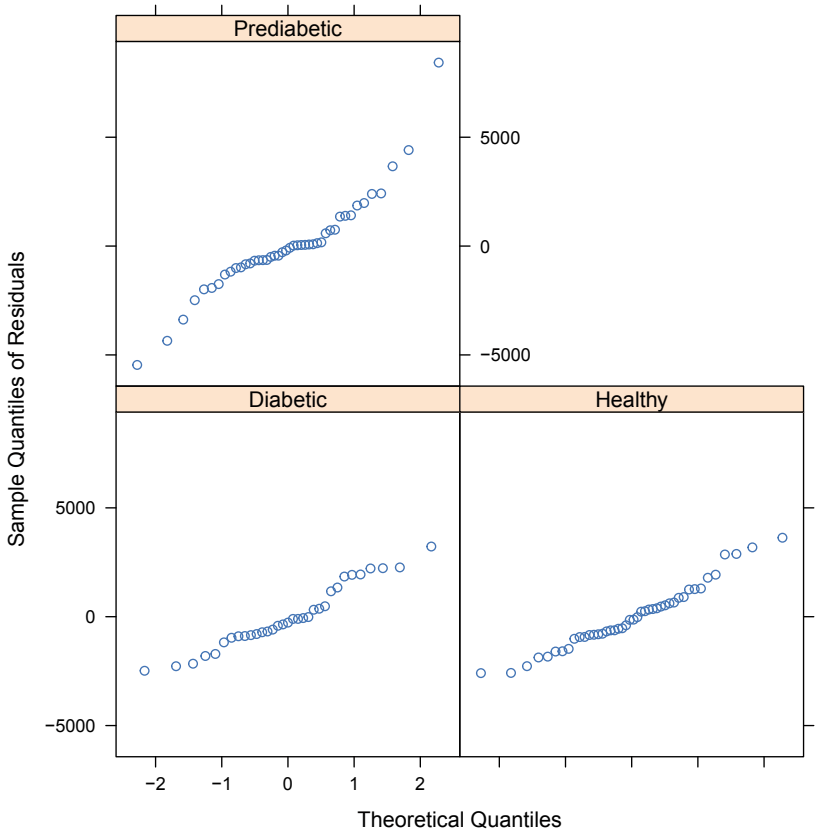
```
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lme4::lmer(formula = Threitol ~ Group + Time..min. + (1 | Participants),
##     data = data)
##
## Linear Hypotheses:
##                             Estimate Std. Error z value Pr(>|z|)
## Healthy - Diabetic == 0        -3730       2481  -1.503   0.2889
## Prediabetic - Diabetic == 0     1206       2481   0.486   0.8778
## Prediabetic - Healthy == 0      4935       2283   2.162   0.0777 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)

res<-glht(mod1,linfct=mcp(Time..min.="Tukey"))
plot(res,main="Threitol")
```




```
summary(res)
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lme4::lmer(formula = Threitol ~ Group + Time..min. + (1 | Participants),
##     data = data)
```

```
##
## Linear Hypotheses:
##              Estimate Std. Error z value Pr(>|z|)
## 30 - 0 == 0   2059.58     487.13   4.228   <1e-04 ***
## 90 - 0 == 0   2132.77     491.61   4.338   <1e-04 ***
## 90 - 30 == 0    73.19     487.13   0.150    0.988
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)

print(qqmath(~resid(mod1)|Group,data))
```


Prediabetic
Diabetic
Healthy
Sample Quantiles of Residuals
Theoretical Quantiles
5000
0
−5000
−2
−1
0
1
2

7

4 FURTHER DEVELOPMENT OF THE SUGAR PROFILING METHOD

4.1 Background

Although the semitargeted one-dimensional GC-MS sugar profiling method already enabled detection and separation of 55 known and unknown sugar compounds in human urine (section 3.2), we still observed further trace signals of compounds that exhibited mass spectra characteristic for sugar compounds. Additionally, in a number of cases the specificity in the one-dimensional GC-MS method is based not on chromatographic resolution, but on the detection of characteristic mass fragments. For targeted approaches with exhaustive sample clean-up the separation via *m/z* is usually sufficient. However, in the case of metabolite profiling approaches with only minimal sample clean-up (as for the semitargeted GC-MS sugar profiling method), the risk exists that unknown matrix compounds and non-sugar compounds with similar mass fragments interfere with the detection of the targeted compounds. Another relevant point with respect to separation of sugar compounds based on MS is that the mass spectra of stereoisomers are highly similar to identical and therefore, separation based on MS is problematic. As a consequence, further optimization of the sugar profiling method with respect to sensitivity and separation performance is desirable. GC×GC-MS enables both an improved sensitivity (via band focusing in the modulator) and an improved separation performance (via separation on two orthogonal columns).

For the transformation of the one-dimensional GC-MS method towards a two-dimensional GC×GC-MS method, several points have to be addressed: i) the column set-up; ii) the MS mode; and iii) the data processing strategy. With respect to the column set-up, the 60 m long Rxi-5SilMS column used in the one-dimensional GC-MS method was chosen as ^{1}D-column and for the ^{2}D-column a classical medium polar BPX50 column with a separation length of 1 m was utilized, leading to a typical orthogonal unpolar × medium polar column combination (introduction section 1.1.3). The one-dimensional GC-MS method is performed in Scan/SIM mode, enabling both detection and elucidation of unknown compounds using Scan mode and the sensitive and selective detection using SIM mode. However, Scan/SIM mode is no longer an option using GC×GC-MS due to the necessity to achieve a fast data acquisition rate for an adequate reconstruction of the narrow GC×GC peaks (introduction section 1.1.3). To improve sensitivity and selectivity, characteristic SIM masses were

considered as the mode of choice for the GC×GC-MS method. The usage of SIM mode for two-dimensional GC×GC-MS is unusual and has only rarely been described so far [88-90]. The characteristic *m/z* for sugar compounds were chosen based on experiences from the one-dimensional GC-MS method (section 3.3 results). Finally, as mentioned in section 1.1.3, processing of two-dimensional data is more complex than in the case of one-dimensional data and necessitates dedicated software and algorithms. In our workgroup, a peak-feature based approach for GC×GC-Scan-MS was established by Björn Egert and Christoph Weinert [31, 91]. This approach so far did not enable the evaluation of SIM data and usage of targeted ions. With the help of Björn Egert the peak-feature based approach was extended towards GC×GC-SIM-MS data. The applicability of the optimized GC×GC-SIM-MS method and the advantages over the so far utilized one-dimensional method are shown exemplarily based on data of a cross-over intervention study (FoodBAll) with multiple urine collections after the consumption of apple, coke, and water as control. The focus of this chapter lies on the further technical development of the sugar profiling method in comparison to the one-dimensional method. Nonetheless, some preliminary impressions on the biological results of the FoodBAll intervention study are presented.

4.2 Material and methods

Chemicals

Methanol (GC; ≥ 99.9%) was purchased from Carl Roth (Karlsruhe, Germany). Heptane (GC) and acetone (GC) were from Merck (Darmstadt, Germany). Pyridine (anhydrous, 99.8%) was supplied by Sigma-Aldrich (Steinheim, Germany). *O*-Methoxylamine hydrochloride was obtained from Chemos (Regenstauf, Germany). *N*-Methyl-*N*-(trimethylsilyl)trifluoroacetamide (MSFTA) with 1% trimethylchlorosilane (TMCS) was from Macherey-Nagel (Düren, Germany). Supplemental Table S4.1 lists all used standard substances.

Consumables

Consumables utilized in this study are listed in Supplemental Table S4.2.

Study design and samples

The samples were generated during the FoodBAll study at the MRI in February 2016. The study was a randomized crossover intervention study with the aim to identify

potential markers of apple and coca cola consumption, including water as a control food and a reference meal (Keto-Drink, Tavarlin GmbH, Pfungstadt, Germany), which was ingested together with the test foods (apple (Elstar), Coca Cola® and water) in three separate arms of the study. For the detailed insight on the study design, see Figure 4.1.

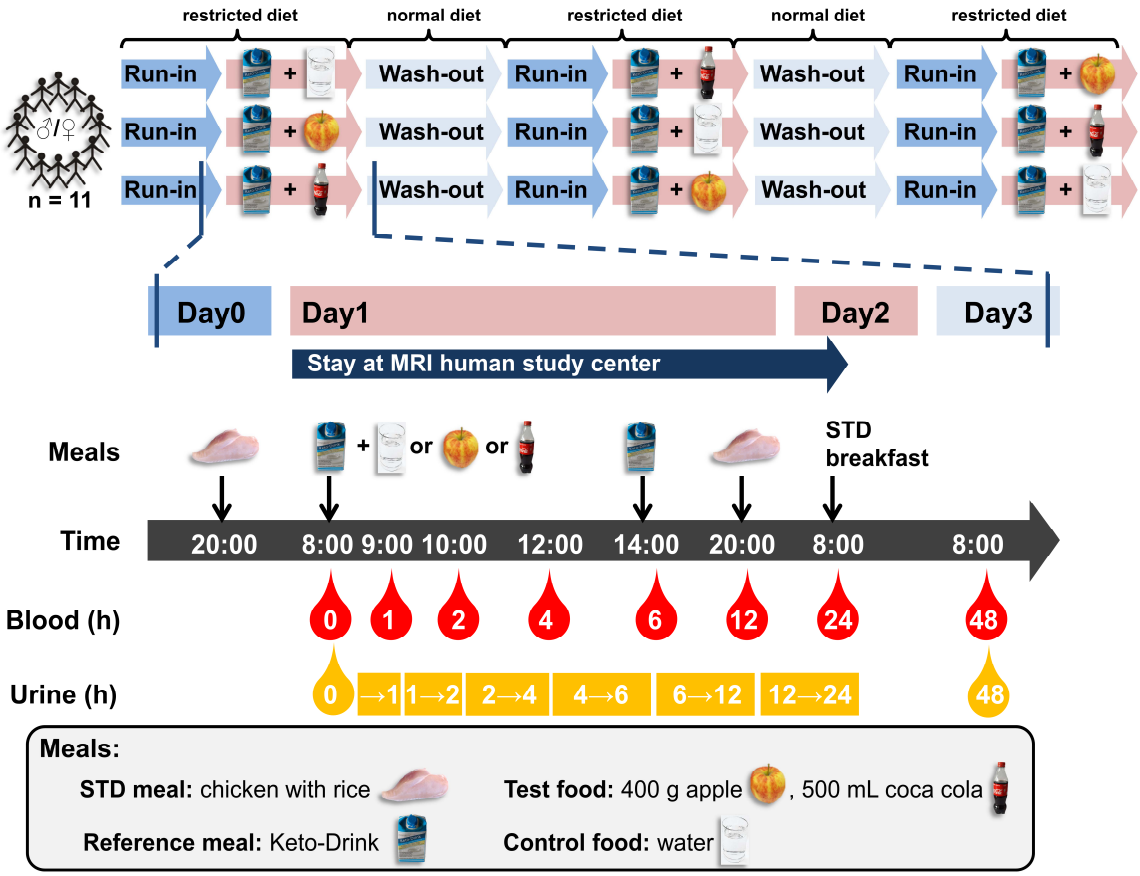


Figure 4.1. Study design of the FoodBAll MRI study.

In total, 11 participants (5 males and 6 females), aged between 21-33 years, BMI 19.7-27.1 kg/m², who gave their written consent, were recruited. Participants were included, if they were healthy male or female, between 18 and 40 years, had a BMI between 18.5 and 30 kg/m², were free from prevalent diseases, had no history of a chronic disease, were non-smokers, did not take any medication (e.g. hormonal contraceptive), or did not take supplements within the last four weeks, did not take antibiotics within the last six months, donated blood within the last three months, in case of women were not pregnant or lactating, did not have an allergy or intolerance for one of the test food or reference meals and were willing and able to perform the study. The ethics committee of the State Medical Chamber of Baden-Württemberg, Stuttgart, Germany, approved the study, which was in accordance with the 1964

Helsinki declaration and its later amendments. The study was registered at the German Clinical Study Register (DRKS00008787) and has the Universal Trial Number U1111-1177-1536.

For optimization and development of the GC×GC-MS sugar profiling method only the urine samples were used. Participants provided spot urine samples before the intake of the test food, at 48 hours after the intervention and collected their urine during the periods described in Figure 4.1. During collection periods the bottles were kept on ice, until samples were processed. After recording urine volume, samples were centrifuged (10 min, 1850 × g, 20 °C), and aliquoted, frozen at -20 °C and after one day stored at -80 °C.

A QC sample was prepared by mixing urine samples from all participants, all interventions and all time points. Separate aliquots were frozen at -80 °C until analysis. Osmolality of the QC sample and study samples was determined in duplicate after centrifugation by using an Advanced® Osmometer Modell 2020 (Advanced Instruments, Massachusetts, USA).

Sample preparation

First, 40 µL of urine sample (study and QC samples) were diluted with water (dH_2O) to adjust osmolality to 60 mOsmol/kg. 20 µL of IS solution (composition see Supplemental Table S4.1) were pipetted to a screw thread vial with a 300 µL fused insert and 40 µL of the diluted urine sample were added. After mixing, the samples were evaporated in a Speedvac-concentrator (Heto maxi dry plus, Jouan Nordic) for 1 h at 40 °C and maximal vacuum ($p < 1$ mbar). To remove remaining traces of water, 20 µL of methanol were added followed by an additional evaporation step (30 min). After evaporation, a two-step derivatization protocol was performed. First, methoximation was conducted by adding 15 µL of methoxylamine-hydrochloride solution (20 mg/mL in anhydrous pyridine) at 40 °C for 1 h under shaking (1000 rpm). Second, trimethylsilylation was carried out by adding 50 µL of MSTFA with 1% TMCS at 75 °C for 1.5 h. Each day, a solvent blank was prepared similarly to urine samples, but with water instead of urine. At the end of the reaction, 10 µL of the FAME solution for RI were added (Supplemental Table S4.1). For the sugar compound reference standard samples (Supplemental Table S4.1) measured at the end of the measurement series, 50 µL of the 25 µMol/L standard solutions were evaporated to dryness in a two-step procedure as described before. Derivatization conditions were

the same as for study samples, however 10 µL methoxylamine-hydrochloride solution and 30 µL MSTFA with 1% TMCS were used.

GC×GC-MS analysis

To correct drift and offset effects during the measurement due to shifts in equipment performance, analyte degradation and also prolonged derivatization reactions or changes in the system (e.g. change of septum), the regular measurement of QC samples is necessary, enabling an analyte-wise correction of these effects. Therefore, each day started with the daily solvent blank sample followed by 11 study samples in three groups of four/three samples framed by six QC samples, with two in the beginning and in the end. In the beginning and in the end of each measurement week, a QC sample was measured in Scan mode (*m/z* 60-550), this run can later be used for identification based on the full scan. At the end of the measurement series, reference standards of sugar compounds (Supplemental Table S4.1) were measured, so that they could be aligned with the study and QC samples and thus enable an *a priori* identification of known sugar compounds. Every measurement week, the liner and injector septum were replaced and the qMS was tuned, followed by six equilibration runs with QC samples. Table 4.1 lists the GC×GC-MS system and software utilized in this study. In Table 4.2, the GC, modulation and MS parameters are shown.

Table 4.1. GC×GC-MS system and software.

Instrument component/ software	**Name**	**Manufacturer**
Gas chromatograph	GC-2010	Shimadzu Corp, Kyoto, Japan
Mass spectrometer	QP2010 Ultra	Shimadzu Corp, Kyoto, Japan
Autosampler	AOC-20i+s	Shimadzu Corp, Kyoto, Japan
PTV Injector	OPTIC-4	GL Sciences, Eindhoven, The Netherlands
Modulator	ZX2	ZOEX Corp., Houston, USA
GCMS instrument software	GCMSSolution 4.11	Shimadzu Corp, Kyoto, Japan
PTV software	Evolution Workstation 4.1	GL Sciences, Eindhoven, The Netherlands
GC×GC visualization software	ChromSquare 2.1	Chromaleont Srl, Messina, Italy
FiehnLib database	FiehnLib	Agilent, Santa Clara, USA
NIST database	NIST2014	NIST, Gaithersburg, MD, USA

Table 4.2. Method parameters including consumables.

Parameter	Setting/value
Autosampler parameters	
Syringe	10 µL, conical
Injection volume	1.2 µL
Pre Clean with Solvent (heptane, acetone)	3
Post Clean with Solvent (heptane, acetone)	6
Pre Rinses with Sample	1
Plunger Speed (Suction)	High
Viscosity Comp. Time	0.2 s
Plunger Speed (Injection)	High
Syringe Insertion Speed	High
Pumping Times	3
Injection Port Dwell Time	0.3 s
GC parameters	
Carrier gas	Helium
GC mode	Constant velocity
Initial column head pressure	210.3 kPa
Liner type	Deactivated split liner with quartz wool (CS chromatography, Langerwehe, Germany)
Injector septum	Septa Thermolite® Shimadzu Plug (Restek, Bad Homburg, Germany)
^{1}D-column	Rxi-5SilMS; ^{1}L = 60 m plus 10 m of an integrated pre-column; $^{1}d_c$ = 0.25 mm; $^{1}d_f$ = 0.25 µm (Restek, Bad Homburg, Germany)
^{2}D-column	BPX50; $^{2}L_{total}$ = 2.4 m, including a “separation segment” of $^{2}L_{sep}$ = 1.0 m; $^{2}d_c$ = 0.15 mm, $^{2}d_f$ = 0.15 µm (SGE, Milton Keynes, United Kingdom)
Column connector	SilTite MiniUnion (SGE, Milton Keynes, United Kingdom)
GC temperature ramp	80 °C → 8 °C/min → 140 °C → 1.75 °C/min → 220 °C → 8 °C/min → 255 °C → 4 °C/min → 300 °C (6.16 min). Run time: 75 min.
Injection mode	Cold split
Split ratio	1:5 → 1:20 (1 min)
PTV temperature ramp	90 °C → 60 °C/s → 280 °C, hold until end of run.
Interface temperature	280 °C

Table 4.2 continued.

Modulation parameters	
Modulator type	Cryogenic, air-based, loop-type
Modulation period (P_M)	2.4 s
Cold jet temperature	-90 °C
Hot jet temperature	Programmed stepwise, at least 100 °C above oven temperature
Hot jet duration	200 ms
MS parameters	
Ion source temperature	200 °C
Ionization mode	EI (70 eV)
MS Mode	SIM
Event time	30 ms
SIM *m/z*	160.00, 277.00, 292.00, 305.00, 307.00, 318.00, 319.00, 333.00, 361.00, 421.00, 437.00
Data acquisition frequency	33 s^{-1}

Data processing

Data processing is a complex and labor-intensive process involving different steps and different softwares (GCMSSolution, R modules, Microsoft Excel). The original workflow for a peaklet-based generic analysis of untargeted GC×GC-MS data was developed by Björn Egert and Christoph Weinert [31, 91]. This strategy is based on full scan data. However, for the sugar profiling, the measurement was performed in SIM mode with 11 selected *m/z*. To address the differences between full scan and SIM data, the workflow was adjusted with the help of Björn Egert. In the following paragraphs, a short description of the relevant points is given, a flow diagram can be found in Supplemental Figure S4.1:

- **Raw data processing:**

  As a first step, raw data was submitted to automated integration. For this, overall 44 integration methods were created, four for each measured *m/z* (integration parameters see Supplemental Table S4.3). This was necessary due to the limitation of the GCMSSolution software (designed for the processing of one-dimensional GC-MS data) being able to handle only a maximum of 1000 peaks per run. Therefore, the chromatogram had to be integrated within timeframes for each of the 11 measured *m/z* with an overlap of 1 min between the time frames (Supplemental Figure S4.1, part A). A batch

process resulted in one txt-file per run (overall 470 runs: QC samples, study samples, technical replicates and standard substances) including the 44 integration methods. Each retention time can be reported up to 11 times depending on the number and size of peaks for each of the *m/z*.

- **Import and restructuring of data:**
  Each txt-file consisted of peak lists for four time frames of one *m/z* followed by the time frames of the next *m/z* and so on in a block-wise structure; information such as retention times (apex, start and end time), area, height and RI are included. Additionally, after each time frame block, the mass spectra consisting of maximally 11 masses for each listed peak are reported (Supplemental Figure S4.1, part B). The txt-files were imported and reorganized towards a tabular structure. Therefore, the 11 *m/z* were assigned to their respective block peak list based on a heuristical approach. The overlapping time frames of 1 min were removed based on exactly the same retention time and mass spectra. Additionally, based on the modulation time the retention time (RT) of the raw data was converted to ^{1}D- and ^{2}D-RT's. To compensate RT-shifts for the modulations (Supplemental Figure S4.1, part C), runs were shifted batch-wise for maximally two modulations to the left or right. Which batches or measurement days were shifted and in which direction was decided based on the median RT's for the highest modulation of three randomly chosen sugar compounds in the QC samples of each day. The result of this step was a first global data matrix with a continuous peak list for each *m/z* and run.
- **Filtering and data reduction:**
  In the next step, a data reduction was performed based on several filters. All peaks with peak heights below 5000, and less than five data points were deleted. Based on the 2D-chromatogram obvious noise peaks and background peaks were deleted by adding their mass spectra to a noise database. Additionally, all peaks were deleted that could not find a matching peak (RT and mass spectral similarity) in any other run.
- **Alignment:**
  The alignment is performed to combine all modulations in one run and to combine matching peaks over all runs. In a first step, a clustering algorithm combined peaks based on their ^{1}D- and ^{2}D-retention times and mass spectral

similarity ($\varepsilon < 0.015$) by comparing peaks pair-wise (Supplemental Figure S4.1, part D). In a second step, each resulting cluster was evaluated again by calculating a hierarchical cluster analysis of the mass spectra within each cluster. If the height cut (HC)-value was above 275, the cluster was cut into two separate clusters (Supplemental Figure S4.1, part E). These steps were first performed for six QC samples and nine study samples comprising all study groups leading to a representative reference alignment. Such a reference alignment was necessary to cope with the huge amount of data due to limits in main memory for the computational calculation. Therefore, the processing of large cohorts with many samples is enabled. Then for each peak in each run, the mass spectral similarity was calculated against this reference alignment. The third and last step was the separation of clusters based on their peak height profile. Here, modulations of two separate peaks were combined due to their mass spectral similarity although being two separate peaks (Supplemental Figure S4.1, part F). To find the optimal split between such peaks, the cumulative heights of all runs were calculated. If adding a modulation results in an increase of less than 1.5%, this modulation was assigned to the first of the two peaks and all other modulations build their own cluster.

- **Demodulation:**

  First, the quantifying ion was determined, which is the *m/z* with the highest signal intensity in most modulations of a cluster. Then, for each run, the signal intensities of the modulations of the quantifying ion were summed up to receive one signal intensity for one analyte in each run.

- **Drift correction:**

  As a result of long measurement series, drift or batch effects can occur due to increasing contamination of the injector and column by matrix compounds, due to batch-wise derivatization (e.g. instability of derivatives, prolonged derivatization) or due to maintenance (e.g. changing of liner or septum). Representative QC samples were injected six times per batch/measurement day to correct such effects. Therefore, analyte-specific correction functions were calculated based on QC samples, and the signal intensities of study samples were adjusted accordingly (Supplemental Figure S4.1, part G). In this step, analytes were separated in correctable and uncorrectable analytes. If an

analyte was found with less than 15% of the QC samples per batch/measurement day in more than 50% of all batches/measurement days, then the analyte was defined as uncorrectable. No correction of drift effect was performed because neither a batch-wise nor an analyte-specific correction function is possible with so few QC samples. For measurement days/batches with less than 15% QC samples or where the first and/or last QC samples were missing, only a global batch correction via the mean of the QC samples was performed. For all other batches/measurement days, where the frequency of the analyte in the QC samples was high enough, analyte-specific correction functions were calculated. The result was a final data matrix for each correctable and uncorrectable analytes, where analytes were listed in columns and samples in rows.

- **Identification:**

  For identification purposes two steps were performed. First, at the beginning and at the end of each measurement week QC samples were measured in Scan mode. Thus, it was possible to perform a comparison of the mass spectrum of a compound with mass spectral libraries (in-house database, NIST14, FiehnLib). Second, overall 111 sugar reference standards were measured in 35 runs at the end of the measurement series (Supplemental Table S4.1). Therefore, the alignment of these reference standards with the QC and study samples and thus the identification of known sugar compounds was enabled.

- **Evaluation:**

  After the data processing, an evaluation of the analytes had to be performed with respect to the choice of quantifying ion, with respect to a correct alignment and height profile split, with respect to the repeatability and with respect to the occurrence of partial or complete coelution with matrix compounds, non-sugars or other sugar compounds.

4.3 Results and Discussion

The optimized two-dimensional GC×GC-MS method enabled the detection and separation of 52 known sugar compounds, all identified by reference compounds. Furthermore, 28 sugar compounds that could be tentatively identified (e.g. as deoxypentitol or deoxysugar acid) or classified as one of the subclasses of sugar

compounds (e.g. polyol, sugar acid or similar) were detected. Additionally, further compounds were detected that exhibited typical mass spectra of sugar compounds (also section 3.2-3.3), of which only those were evaluated that showed biologically interesting time courses during the intervention with apple and coke or were detected also with the one-dimensional GC-MS method (4 compounds). It is to be expected that about 10-20 further unknown sugar compounds might be detectable in urine using the two-dimensional GC×GC-MS method based on the mass spectra. Overall, 84 known and unknown sugar compounds were detected. In panel A of Figure 4.2, a typical two-dimensional GC×GC-MS chromatogram of a urine QC sample is displayed exemplarily.

The tentative identification and/or classification of 28 as yet unknown sugar compounds was possible as a result of the mass spectra and also of the so-called *structured retention* effect, which results from the separation of compounds on two columns with independent/orthogonal separation mechanisms. In addition to the ^{1}D-separation based on volatility, and thus more or less on molecular weight, compounds were separated on the ^{2}D-column based on additional polar interactions. Consequently, the number of C-atoms and the subclass (i.e. polyols, monosaccharides, sugar acids or sugar acid lactones; Figure 4.2 panel B) of an unknown sugar compound can be deduced based on the chromatographic position in the 2D-chromatogram, especially in conjunction with the mass spectra.

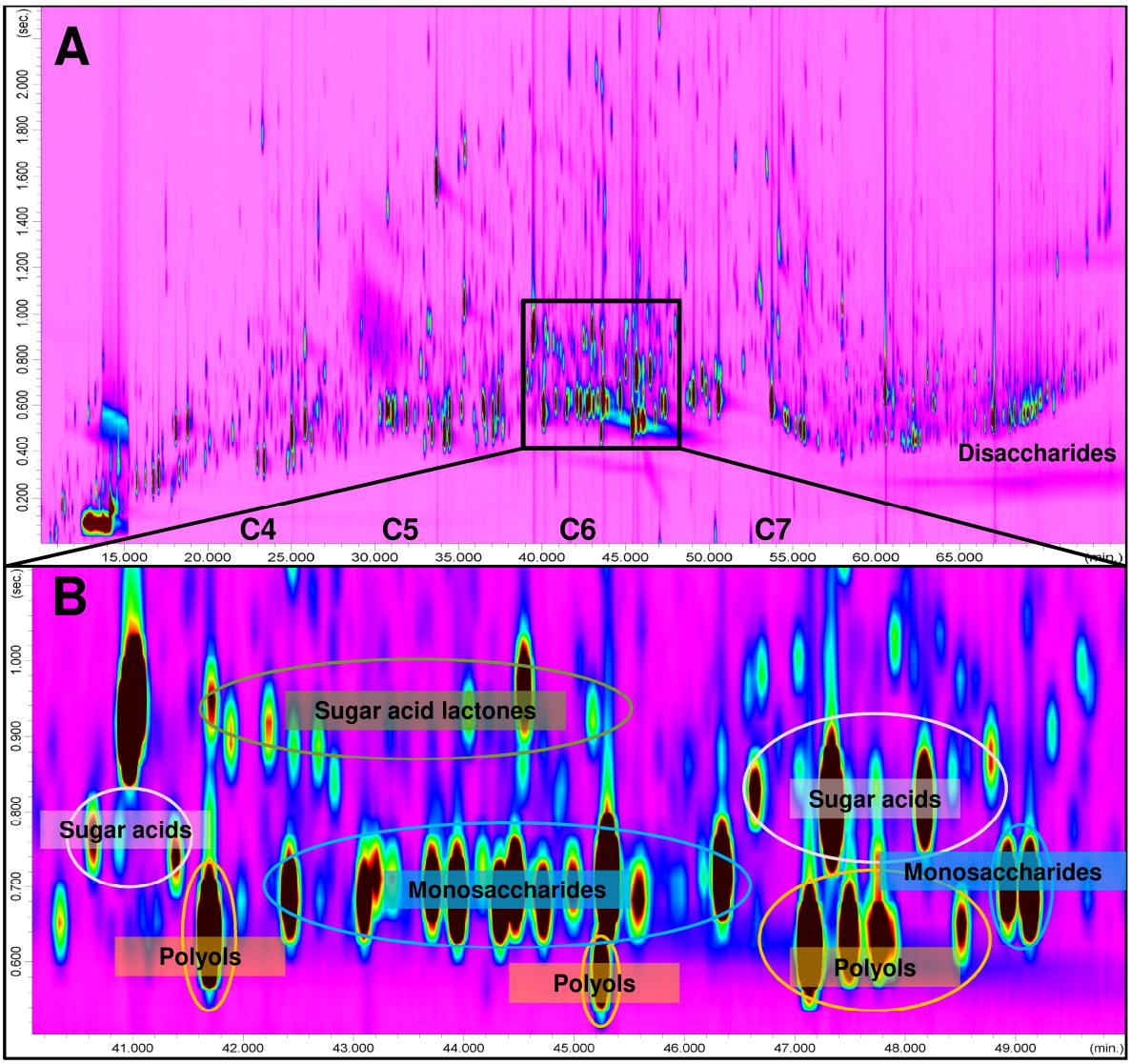


Figure 4.2. A: GC×GC-SIM-MS chromatogram of a urine QC sample. C4-C7 marks the regions, where tetroses, pentoses, hexoses and heptoses elute. B: Image section including chromatographic regions for sugar acid lactones, sugar acids, monosaccharides, polyols and disaccharides.

In the Supplemental Tables S4.4-S4.8, the sugar compounds detected in human urine with the one-dimensional GC-MS method and the two-dimensional GC×GC-MS method are listed separately for the subclasses monosaccharides, disaccharides, polyols, sugar acids and others/unknowns. The known sugar compounds from the one-dimensional GC-MS sugar profiling were taken from Supplemental Table 3 of Publication No. 2 (section 3.2). In general, due to the improved sensitivity and separation performance a higher number of sugar compounds in all subclasses could be detected using the GC×GC-MS method in comparison to the GC-MS method (Figure 4.3). All the known sugar compounds and most of the unknown ones found

with the one-dimensional GC-MS method were also detected with the GC×GC-MS method, except for some unknown sugar compounds. Only two compounds (ribonic and xylonic acid) were separated using the one-dimensional GC-MS method, but coeluted using the two-dimensional method.

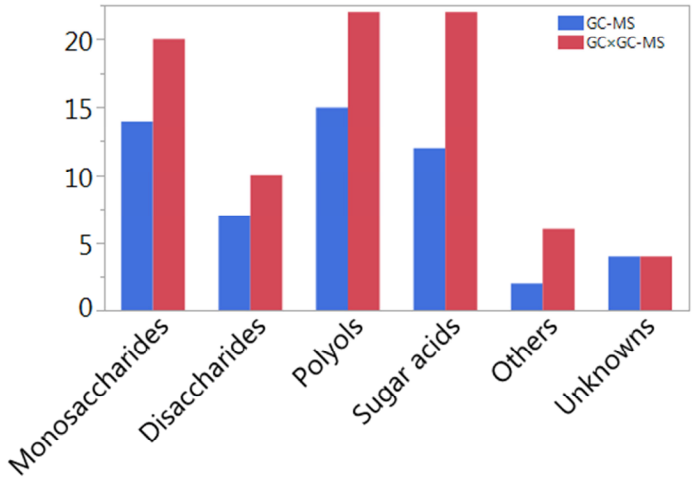


Figure 4.3. Number of detected sugar compounds in each subclass of sugar compounds for the GC×GC-MS and the GC-MS method.

The improved sensitivity is a result of both the usage of GC×GC-MS and the utilization of characteristic mass fragments of sugar compounds. As a result of the band focusing in the modulator before the short separation on the ^{2}D-column, the analytes reach the detector as very narrow peaks (50-200 ms peak base width), and therefore the signal-to-noise ratio and consequently sensitivity is improved (section 1.1.3). Another factor playing a role for the increased sensitivity of two-dimensional GC×GC-MS is that chemical or chromatographic noise (result of column bleed and/or solvent tail) is separated as a compressed band from most of the analytes [92]. Typically, the sensitivity is increased by a factor of 10, which has been discussed exhaustively in the literature [92-94].

Next to the GC×GC method in general, the usage of the MS in SIM mode enables an improved sensitivity through longer dwell times for each selected mass. Consequently, a larger number of selected ions reach the electron multiplier and the signal intensity, and thus signal-to-noise (S/N) ratio, is increased. To highlight the difference in sensitivity between scan and SIM analysis, the mean S/N ratio was calculated for the major modulation of 15 selected sugar compounds measured in four urine QC samples each. As shown in Table 4.3, sensitivity was on average 3.2 times higher in SIM mode, compared to Scan mode. Figure 4.4 depicts an

overlay of chromatograms of a urine QC sample, which have been recorded in Scan and SIM mode while keeping all other parameters constant. The signal intensity of the sugar compounds was clearly higher in SIM mode compared to the intensity of the same mass traces extracted from Scan data. Therefore, an overall increased sensitivity (S/N) by a factor of approximately 32 can be assumed by combining the increased sensitivity through the GC×GC method (factor 10) and the SIM mode (factor 3.2). Although SIM mode enables an increased sensitivity, there are some disadvantages or problems connected to using only a limited number of mass fragments. First, identification of compounds is complicated and necessitates measurement of reference substances and/or measurement of additional samples in Scan mode for identification purposes. Both was performed for this study and proved useful in identifying and classifying 84 sugar compounds, although the measurement of overall 111 reference substances in 35 runs was somewhat costly in terms of measurement time. Another possibility for the future might be a mass spectral database of sugar compounds measured with the selected characteristic 11 mass fragments, enabling a direct annotation during the processing of data, together with RI's.

Table 4.3. Comparison of S/N ratio of the highest modulation of 15 selected sugar compounds in four urine QC samples measured in SIM mode and four urine QC samples measured in Scan mode, while keeping all other parameters constant.

Sugar compounds	S/N ratio SIM mode				S/N ratio Scan mode				Average S/N ratio		S/N ratio factor SIM/Scan
	QC001a	QC016b	QC017a	QC036b	Scan QC01	Scan QC02	Scan QC03	Scan QC04	SIM	Scan	
Arabinose	2127	1126	718	1460	245	262	447	186	1358	285	4.8
Fucose deriv. 2	193	173	192	1088	50	93	156	81	411	95	4.3
Fructose deriv. 2	1714	2144	737	874	462	795	657	276	1367	547	2.5
Glucose deriv. 1	3299	3656	984	1952	1458	1683	1822	1990	2473	1738	1.4
Sedoheptulose deriv. 1	5794	3393	2641	3525	175	1056	564	444	3838	560	6.9
Sucrose	1602	1437	789	503	973	445	96	130	1083	411	2.6
Lactose deriv. 1	469	344	776	496	62	55	73	260	521	112	4.6
Threitol	2306	3402	1188	2768	246	869	2372	435	2416	981	2.5
Ribitol	2920	4203	1357	2336	662	1783	687	327	2704	865	3.1
Sorbitol	949	897	789	1510	306	248	440	386	1036	345	3.0
scyllo-Inositol	1354	2153	896	2500	1193	555	963	928	1726	910	1.9
Erythronic acid	1576	2782	1155	1078	1080	1550	2622	910	1648	1540	1.1
Arabonic acid	1229	1388	754	3361	366	664	556	579	1683	542	3.1
Glucuronic acid deriv. 1	1346	2277	888	3652	508	594	1277	713	2041	773	2.6
Levoglucosan	855	691	656	828	66	216	348	264	758	223	3.4
Average of S/N ratio factor between SIM/Scan for selected sugar compounds											3.2

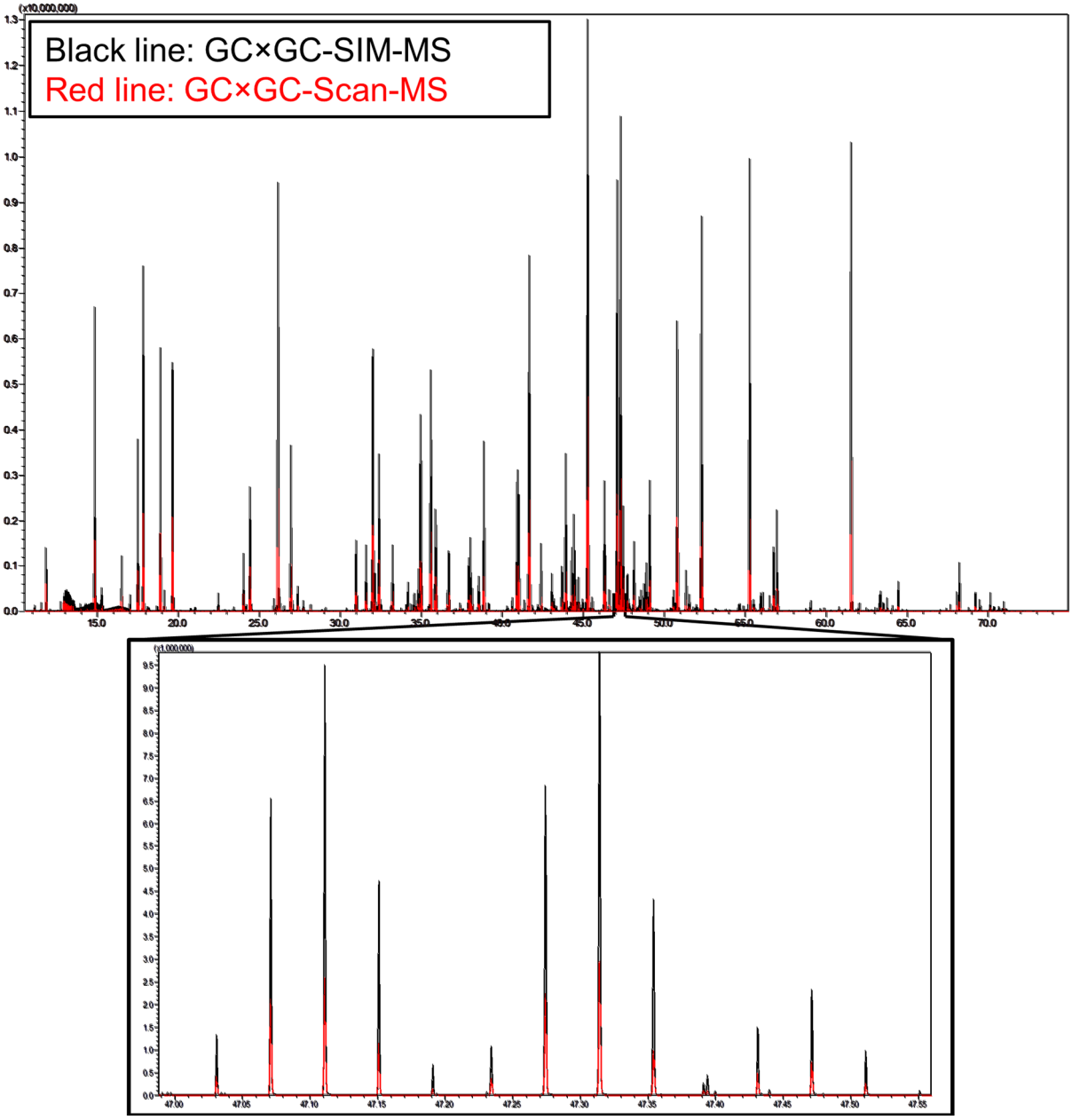
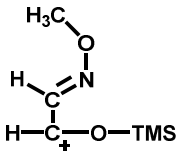
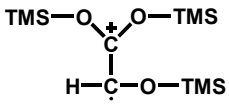
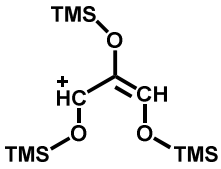
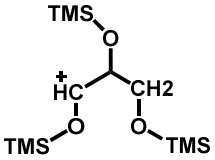
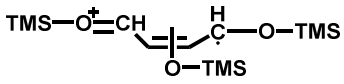
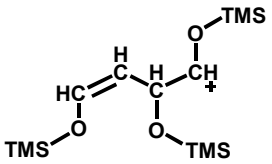
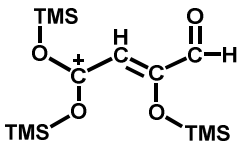
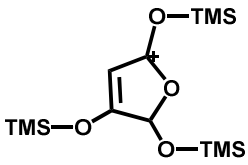
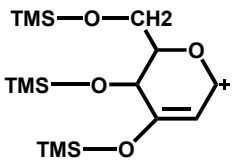
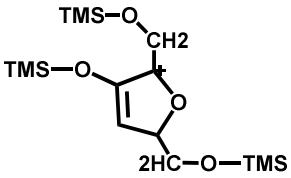
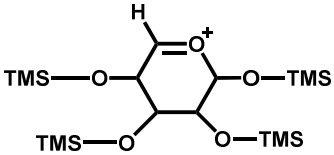


Figure 4.4. Overlay of GC×GC-MS raw data chromatograms of a urine QC sample measured in Scan mode (red) and SIM mode (black). All measurement parameters except the MS mode (Scan/SIM) were the same. For a comparable depiction of Scan mode with SIM mode an extraction of the same 11 *m/z* used for SIM was utilized. Below a section of the chromatogram is increased depicting three sugar compounds (mannitol, glucuronic acid and sorbitol; from left to right).

Second, and maybe most critical, is the choice of the fragments for SIM analysis. As a consequence of the set event time of 0.03 s and the acquisition speed of the qMS instrument, only 11 mass fragments can be utilized for the GC×GC-SIM-MS measurement. The event time was chosen to achieve similar conditions as for the GC×GC-Scan-MS workflow for data processing, which was previously established in

our lab. The 11 characteristic mass fragments were chosen based on the knowledge gained from the semitargeted GC-MS (Scan/SIM) sugar profiling method (sections 3.2-3.3). Some characteristic masses such as *m/z* 204 or 205 were not feasible, because their usage would have led to interferences with non-sugar compounds and consequently a decreased separation performance. In the end, the following *m/z* were chosen: 160, 277, 292, 305, 307, 318, 319, 333, 361, 421 and 437 (Table 4.3). However, the chosen 11 mass fragments define the coverage of sugar compounds detected in human urine. Rarer sugar compounds, which were not in the beginning expected to be detected in human urine, such as 2-deoxyribonic acid-1,4-lactone, exhibited none of the chosen 11 masses in their mass spectra, and therefore could not be detected using SIM mode. While in case of 2-deoxyribonic acid-1,4-lactone, it was possible to detect the corresponding free acid, something similar might not be the case for other unusual sugar compounds. One such example is methyl-*α*-glucopyranoside, which was detected in traces in the QC samples measured in Scan mode, but not in SIM mode. Most of the masses were used and necessary as quantifying ion for at least some compounds. However, both *m/z* 421 and 437 were not used as quantifying ions, because for all detected sugar compounds other more intense fragments were available. Therefore, in future they could be exchanged for other masses, e.g. to detect as yet not covered sugar compounds like 2-deoxyribonic acid-1,4-lactone or methyl-*α*-glucopyranoside.

Table 4.3. Selected mass fragments characteristic for sugar compounds [12].

Mass fragment	Characteristic for	Structure
m/z 160	C2 fragment of aldoses, absent in 2-ketohexoses	
m/z 292	sugar acids with 2,3-dihydroxy structure, formed by rearrangement	
m/z 305	inositols	
m/z 307	sugars, especially C6 ketoses and C5 sugars	
m/z 318	inositols	
m/z 319	sugars	
m/z 333	sugar acids, often combined with *m/z* 292	 
m/z 361	disaccharides and glycosides	 
m/z 437	disaccharides, non-methoximated TMS derivatives of aldoses	

Third, an important aspect in relation with the interpretation of the so far relative quantitative SIM data, is that no comparison of amounts between sugar compounds can be made, because not only response factors of the detector are different (as is the case for normal Scan data), but also the heights of mass fragments are not

comparable for different sugar compounds. Therefore, as is typically the case for any metabolomics experiment with relative quantitative data, the signal intensities of each sugar compound are only comparable within different study groups. This problem can only be avoided by optimizing the method towards an absolute quantification. However, due to the high number of sugar compounds, including even unknown sugar compounds, their broad concentration range, the differential influence of the background matrix and the limited number of isotopically labeled standard compounds, the development of a quantitative method is complex. The only reasonable possibility might be to optimize the method for quantification of a selected number of well-known and relevant sugar compounds. However, the relevance of sugar compounds is highly dependent on the research questions. Therefore, sugar compounds for quantification have to be chosen with respect to the aim of an investigation. In the end, it has to be stated that the semitargeted sugar profiling method has to be viewed as a hypothesis-generating metabolite profiling approach, where only in a second step a quantitative method would be developed to prove the hypothesis and gain further biological insight.

Another important reason for the larger number of detected sugar compounds with the GC×GC-MS method in comparison to the GC-MS method is the separation performance. As mentioned in section 1.1.3 the peak capacity of a two-dimensional GC×GC method is increased in a multiplicative way in comparison to one-dimensional GC. Therefore, separation performance is improved considerably; in Figure 4.5, examples for compounds separated in the ^{2}D are depicted. For example, ribitol and fucose derivative 1 would completely coelute, if measured one-dimensional. The same is true in the case of 1-deoxyhexitol and ribonic/xylonic acid, glucono-δ-lactone, fructose derivative 2 and mannose derivative 1 and in case of *neo*-inositol and glucose derivative 1 (Figure 4.5). Additionally, the effect of the structured retention is observable, i.e. polyols eluting early from the ^{2}D, followed by monosaccharides, sugar acids and sugar acid lactones.

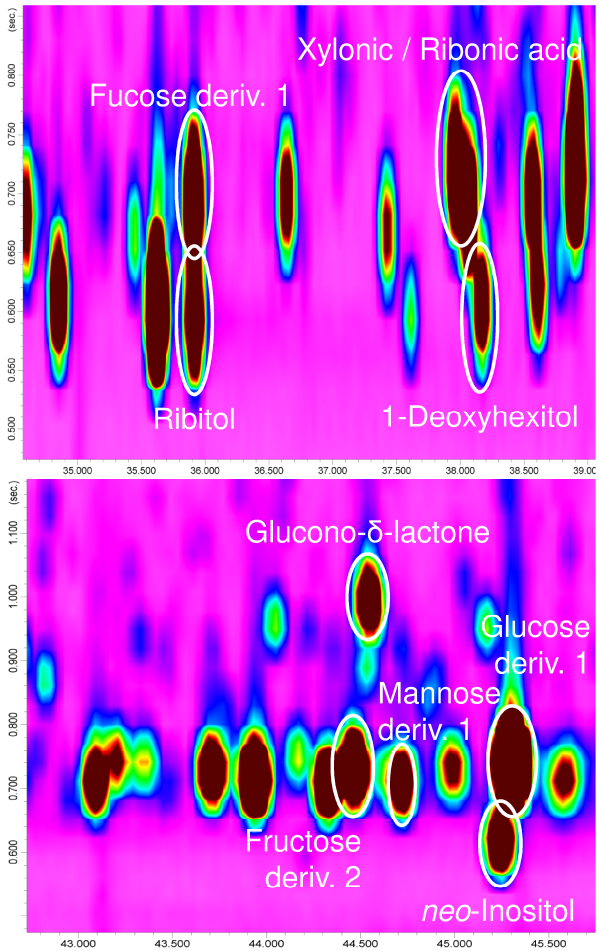


Figure 4.5. The additional value of the ^{2}D-separation. Compounds only separated in the second dimension would coelute in the one-dimensional analysis.

Although separation performance is considerably improved for the two-dimensional method in comparison with the one-dimensional method, in some cases still the usage of the SIM masses for separation from matrix compounds and/or other sugar compounds is necessary (Figure 4.6). However, the choice of characteristic masses enables a sufficient separation from non-sugar compounds and the matrix. While 1-deoxyhexitol A0494_H2_H1 and the unknown sugar acid U09/A0510 coeluted completely, if measured in Scan mode (panel A, 1.-2.), a separation based on the independent mass fragments *m/z* 292 (unknown sugar acid U09/A0510) and *m/z* 319 (1-deoxyhexitol A0494_H2_H1) could be achieved (panel A, 4.). Another example is the separation of psicose derivative 1 from the non-sugar compound quinic acid, which are both partially coeluting in ^{1}D, i.e. at least 1-2 modulations are completely overlapping in Scan mode (panel B, 1.-2.). Nonetheless, using the characteristic mass fragment *m/z* 307 for psicose derivative 1, a complete separation from quinic acid can be achieved (panel B, 4.).

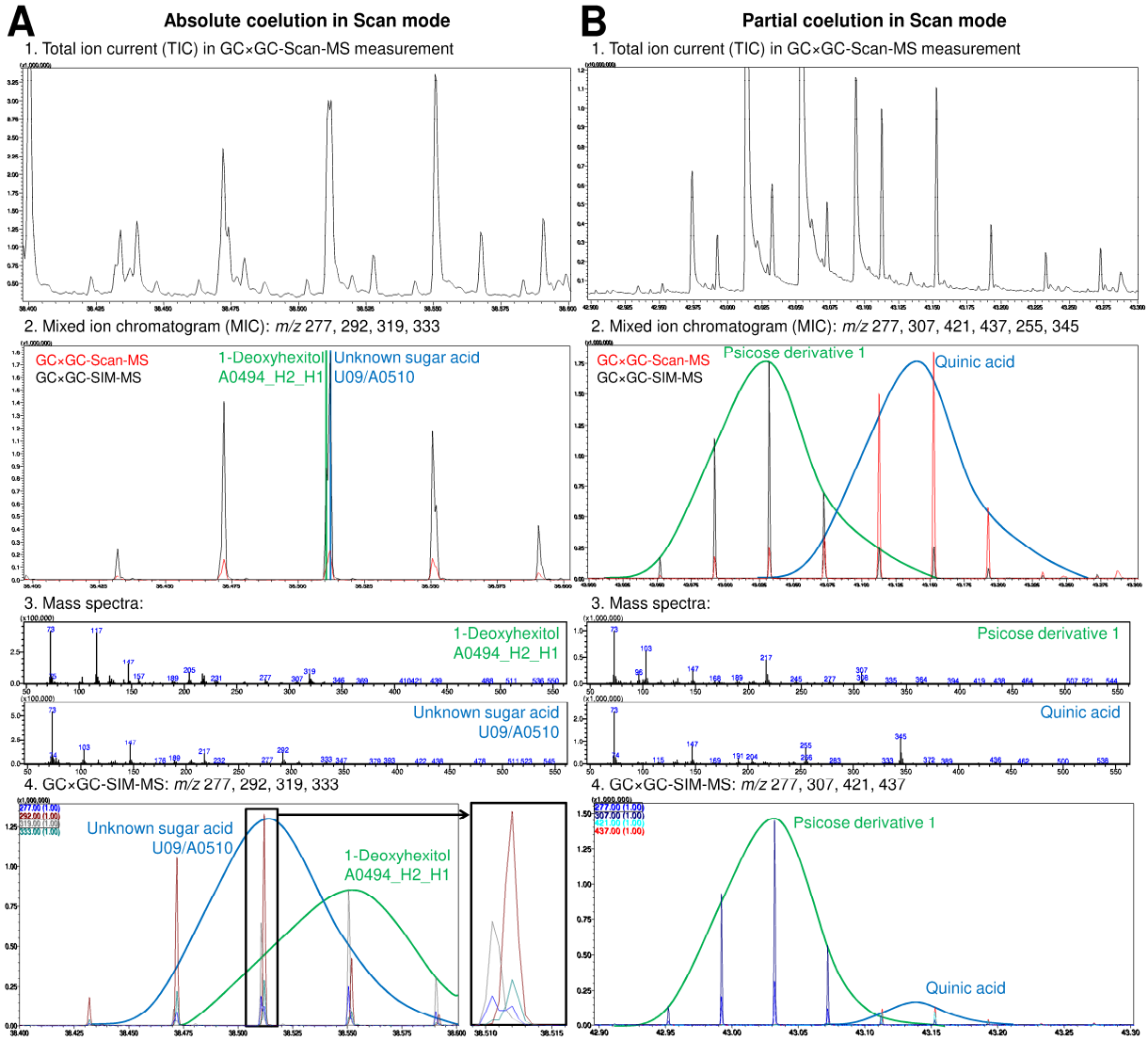


Figure 4.6. Panel A: 1. Absolute coelution of 1-deoxyhexitol A0494_H2_H1 (green line) and unknown sugar acid U09/A0510 (blue line) depicted as TIC chromatogram. 2. MIC chromatogram of the same compounds with the mass fragments *m/z* 277, 292, 319 and 333 measured in Scan (red line) and SIM (black line) mode. 3. Mass spectra measured in Scan mode for 1-deoxyhexitol A0494_H2_H1 and unknown sugar acid U09/A0510. 4. Chromatogram measured in SIM mode, depicting separation via *m/z* 292 and 319. Panel B: 1. Partial coelution of psicose derivative 1 (green line) and quinic acid (blue line) depicted as TIC chromatogram. 2. MIC chromatogram of the same compounds with the mass fragments *m/z* 277, 307, 421, 437, 255 and 345 measured in Scan (red line) and SIM (black line) mode. 3. Mass spectra measured in Scan mode for psicose derivative 1 and quinic acid. 4. Chromatogram measured in SIM mode, depicting separation via *m/z* 307.

Further important aspects for the comparison of the optimized GC×GC-MS method and the GC-MS sugar profiling method are repeatability and robustness during long measurement series. For the one-dimensional GC-MS measurement of the KarMeN

study urine samples, 312 study samples and 144 QC samples were measured. With 264 study samples and 144 QC samples, the GC×GC-MS sugar profiling method is comparable with respect to the number of samples. However, the one-dimensional method has a much shorter run time with 42 min in comparison to 75 min with the two-dimensional method, which resulted in 14 batches for the GC-MS method and 24 batches for the GC×GC-MS method. Retention times for both methods were comparably stable and within an acceptable range. While in the case of the one-dimensional method the peak integration time frames had to be adjusted during data processing, in the case of the two-dimensional method a shifting of retention times (no more than two modulations) was performed to ensure reliable alignment of clusters. In the Supplemental Table S4.9, the inter- and intra-day coefficients of variation (CV's) of all sugar compounds for the GC×GC-MS method are listed together with information on trace and saturated analytes, occurrence of coelution, median heights and frequency in the QC and study samples. The distributions of intra- and inter-day CV's before drift/batch correction for the sugar compounds detected with the two-dimensional GC×GC-MS method is depicted in Figure 4.7. Although the repeatability of the two-dimensional method is slightly worse in comparison to the one-dimensional method (Supplemental Table 4 in section 3.2), it is still sufficient for a semitargeted metabolite profiling approach with intra-day CV's below 15% for 87% of the sugar compounds (Figure 4.7). The slightly better repeatability of the one-dimensional method might be related to the number of data points for each peak: while with the GC-MS method most sugar compounds had about 9-15 data points, in the case of the GC×GC-MS method only 5-9 data points were achieved due to the narrow peaks (Supplemental Table S4.10). For an adequate and precise recording of a peak about 7-10 data points per peak are necessary [28, 33]. While this criterion is easily fulfilled in the one-dimensional method, the data acquisition speed of the qMS does not allow an optimal reconstruction of the narrow two-dimensional peaks. Nonetheless, next to an improved sensitivity and separation performance, a good repeatability over the measurement series with overall 431 samples was achieved for the two-dimensional GC×GC-MS method.

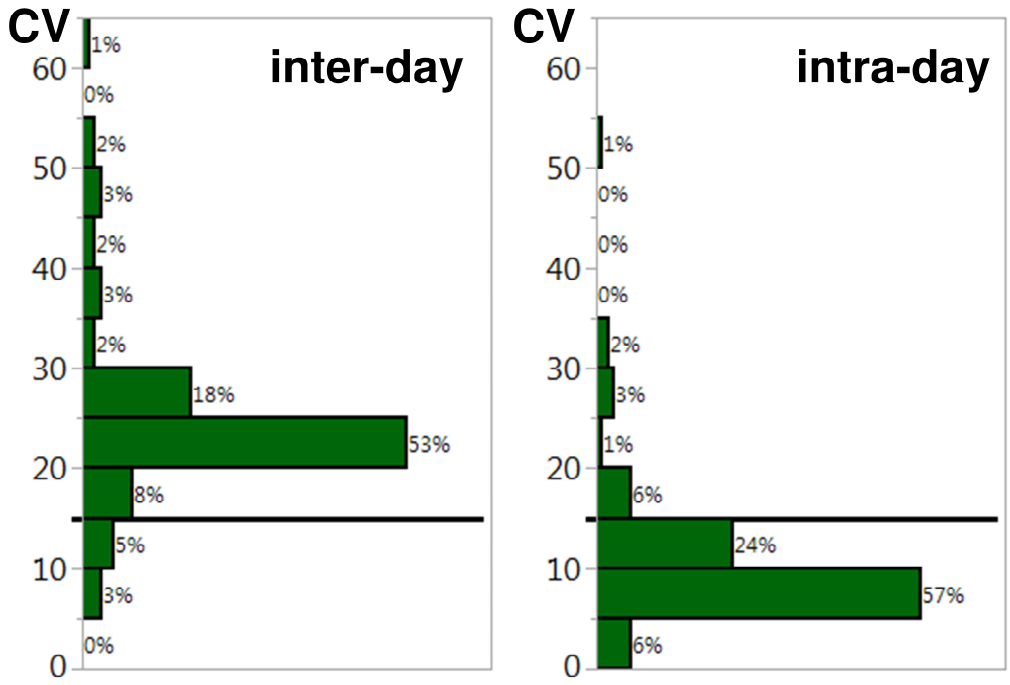


Figure 4.7. Distribution of CV's in QC samples without correction, left: inter-day CV; right: mean intra-day CV.

The data processing for the GC×GC-MS measurement had to be adjusted for the usage of SIM mode. Therefore, the peaklet-based generic analysis of untargeted GC×GC-Scan-MS data [31, 91] was taken as a basis for the development of a SIM-adapted workflow. A number of problems had to be addressed for this optimization:

- One run essentially consists of 11 specific ion mass chromatograms leading to up to 11 mass fragment peaks reported for one modulation (peak). It is clear that this acquisition mode may result in very large data, especially in the case of complex samples.
- The parameters for the alignment and especially the calculation of mass spectral similarity for Scan data were optimized for a *m/z* range of approximately 500 amu, while for SIM-data only 11 *m/z* are available. Therefore, a step-wise adjustment of different parameters (such as mass spectral similarity ε or HC value) in different steps during the alignment is necessary.
- The fact that only 11 fragments are available causes more often the necessity of a height profile splitting due to the highly similar mass spectra of isomeric sugar compounds (Supplemental Figure S4.1, part E). So far, the height

profile splitting is based on the cumulative heights of all runs and results in one split time over all runs. The RT-shifting in modulations during long measurement series can lead to a faulty split for individual samples (Figure 4.8), and this even though a rudimentary shifting of RT was already performed. There are two potential future solutions for this problem, i) a more complex correction of the RT-shifting (including non-linear shifting depending on chromatographic region); or ii) performing the splitting based on each individual run. However, both solutions necessitate complex bioinformatics processes and cannot be easily implemented.

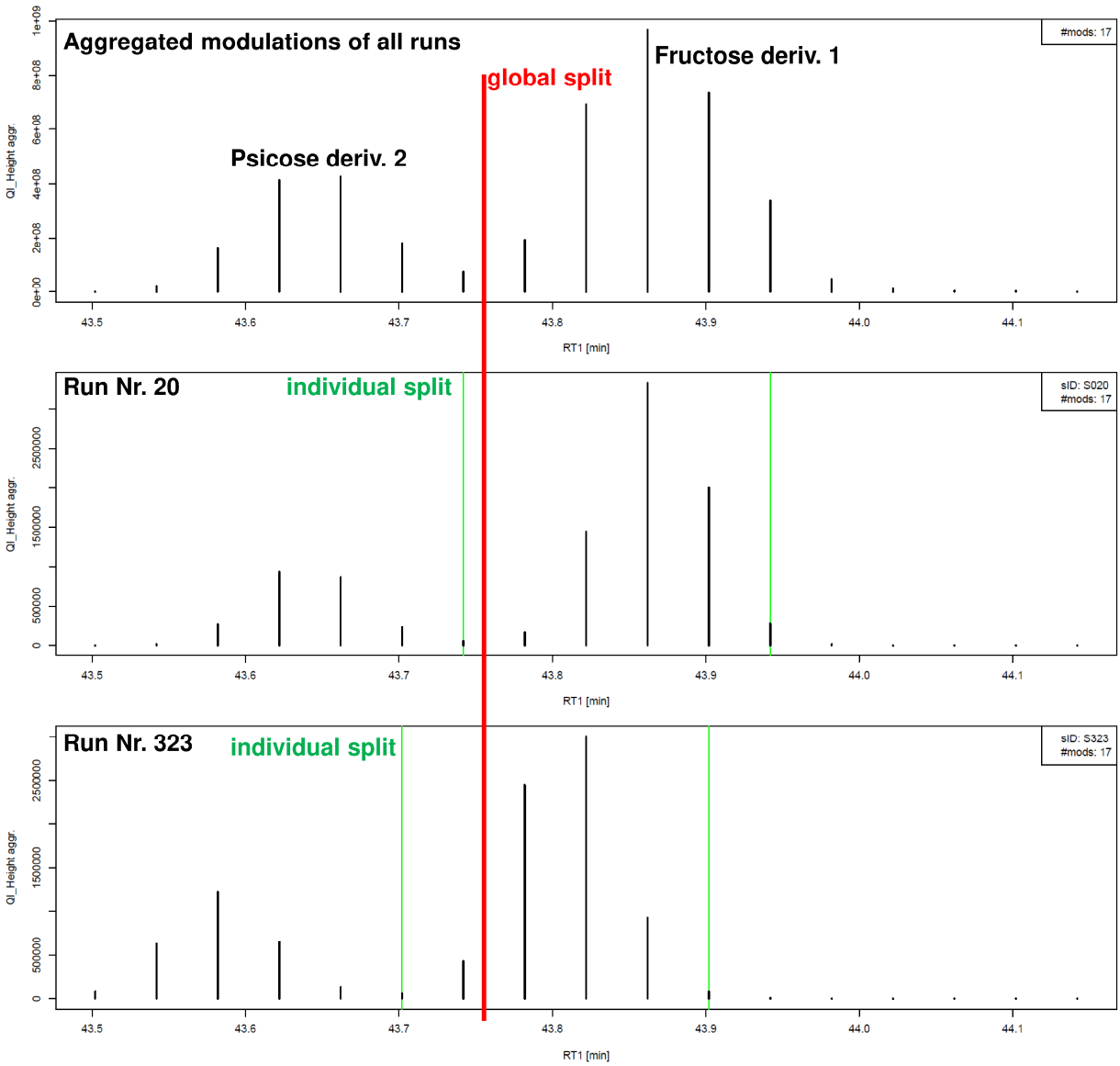


Figure 4.8. Height profile of psicose derivative 2 and fructose derivative 1 for the aggregated modulations of all runs and for two individual runs during the measurement series. Due to their identical mass spectra they were combined in one cluster and had to be splitted using their height profile. In red the global split time is depicted and in green the split time if performed individually for each run.

4.4 Conclusions

The optimization of the so far one-dimensional GC-MS sugar profiling towards a two-dimensional GC×GC-SIM-MS method was successful, and has a number of advantages, the two most important being the improved sensitivity and separation performance leading to the detection of overall 84 sugar compounds. Method repeatability was good with mean intra-day CV's (uncorrected) of less than 15% for 87% of the sugar compounds over a long measurement series with overall 431 samples and 24 measurement days. The GC×GC-SIM-MS sugar profiling method will in future enable the reliable measurement of a large number of sugar compounds in biological samples. Some optimization of the data processing procedure is still necessary. However, good results can already be achieved with the workflow described here. In future, this method can have an important contribution to investigate the metabolism of sugar compounds and will enable the identification of further dietary markers. In Figure 4.9, some first impressions of the biological results of the FoodBAll study are depicted. A number of sugar compounds (e.g. threitol, fructose, sucrose, etc.) increased after the interventions with apple and/or coca cola, but not in the water control. Some of the sugars increase solely after the consumption of the apples and therefore might be promising candidates as markers of apple consumption.

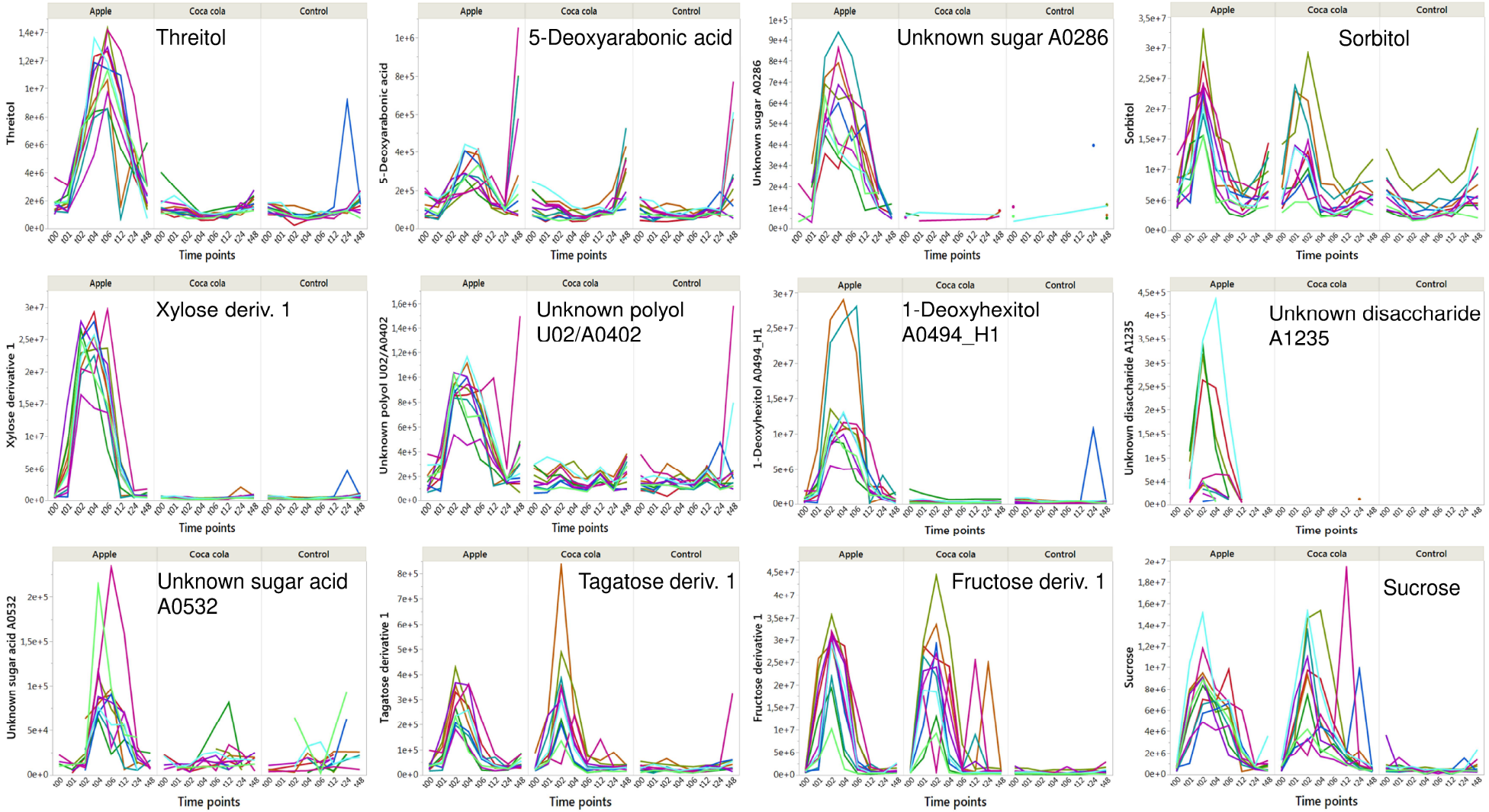


Figure 4.9. Line plots for a number of sugar compounds increasing after apple and/or coca cola intake in urine samples, but not in the control. Each colored line stands for one participant.

5 DISCUSSION AND CONCLUDING REMARKS

Within this thesis, putative markers for dietary intake (sections 3.1, 3.2 and 4) and health status (sections 3.2 and 3.3) were identified using two different GC-MS based metabolite profiling approaches: a semitargeted sugar profiling and a broader, more untargeted VOC analysis. This section aims for an overarching discussion, starting with the classification of the applied GC-MS based metabolite profiling approaches within the broader concept of metabolomics in general. Thereafter, aspects of the discovery of dietary markers are discussed, followed by an outlook with respect to validation of the putative dietary markers identified in this thesis. Last, the application of the semitargeted sugar profiling method towards the discovery of markers for health status, especially with respect to changes in the plasma sugar profile of participants with type 2 diabetes mellitus is addressed.

General classification of the applied GC-MS based metabolite profiling approaches

In general, targeted analysis, metabolite profiling, untargeted metabolomics and metabolite fingerprinting can be differentiated. By definition metabolite profiling approaches, as applied in this thesis, lie in between targeted analysis and untargeted metabolomics approaches. Important aspects for the differentiation of those three approaches are sample preparation, quantification and identification.

While for targeted analysis always an absolute quantification and for untargeted metabolomics so far mostly a relative quantification is performed, in the case of metabolite profiling approaches both relative and absolute quantification is possible. However, more and more scientists demand absolute quantification as the standard for all metabolomics approaches [36, 41], although this will confront the analytical scientists with a number of challenges. These challenges were also the reason that the metabolite profiling approaches developed and utilized in this thesis were of a relative quantitative nature. In general, absolute quantification can be performed using i) calibration curves either in solvent or in matrix; ii) standard addition approach; or iii) isotopically labelled standards (e.g. isotope dilution analysis). Given the sheer number of metabolites, including a huge number of unknown or unusual compounds for which no standards are available, their broad concentration range, the relevant influence of the background matrix on quantification and also the limited, and cost-intensive assortment of commercially available isotope-labeled

standards [95], the effort for the development of an absolute quantification method is enormous and time-consuming. In addition, a number of compromises and potential biases most likely will be included, such as e.g. using an isotopically labelled standard A for a chemically different compound B. A first question here is, whether the advantages of an absolute quantification will outweigh the necessary effort and potential biases and the second question is, whether these biases are comparable to those resulting from a relative quantification. In metabolomics investigations of plant material or microorganisms, another possible, yet cost- and time-intensive, approach is the accurate normalization on a whole isotopically labelled metabolome, which is added as a kind of internal standard. Such an isotopically labelled metabolome is generated by cultivation on fully labelled substrates with ^{13}C, ^{15}N or ^{18}O [96, 97]. However, this is not possible for investigations of human samples, but an interesting, yet also cost- and time-intensive, alternative might be the application of isotopically labelled reagents to create for example an isotopically labelled urine sample [98]. Within the scope of the semitargeted sugar profiling, an absolute quantification would be feasible, especially for selected biologically important sugar compounds, to investigate a defined scientific question. The most important advantage of an absolute quantification in contrast to relative quantification would be that a comparison of amounts between sugar compounds (e.g. glucose and fructose) is possible (section 4).

Identification is one of the other important aspects for differentiation of metabolomics approaches. While in the case of targeted analysis all measured compounds are clearly defined and known, in the case of metabolite profiling approaches additionally to *a priori* known compounds, unknown compounds belonging to the targeted substance class or selected pathway are included (see semitargeted sugar profiling approach, sections 3.2-3.3, 4). In contrast, for untargeted metabolomics, identification is most often performed after statistical analysis for those compounds showing e.g. significant p-values. In this respect, and with respect to the range of chemically diverse substance classes belonging to VOC's, the applied VOC method (section 3.1) might on the one hand be considered an untargeted metabolomics (volatilomics) analysis. On the other hand, in contrast to an untargeted metabolomics analysis, the sample preparation was specific for volatile compounds and as a result limited to a subclass of metabolites showing a certain characteristic or property (in

this case volatility). Therefore, the VOC analysis might also be considered a metabolite profiling approach.

Identification is one of the most essential aspects in any metabolomics approach and at the same time one of the major challenges [12]. Without identification, no biological interpretation or mechanistic elucidation is possible. To ensure proper identification, the Metabolomics Standard Initiative (MSI) proposed a classification of metabolite identification into four levels ranging from unknown compounds to absolutely identified compounds [99]. The most important resources for identification in GC-MS based metabolomics approaches are comprehensive EI-MS based spectral databases, such as NIST, Fiehn Lib, Golm Metabolom Database (GMD) or Wiley database, which include both mass spectral and RI information (based on alkanes or FAME) for library matching of derivatized and underivatized compounds [2, 12, 100]. Those can successfully be utilized across different laboratories due to highly reproducible mass spectra with a standardized ionization energy of 70 eV [100, 101]. In this thesis, identification of volatile marker candidates was performed using such EI-MS based spectral databases with mass spectral and RI information (section 3.1). However, library matching and spectral similarity can lead to ambiguous identification:

- due to problems resulting from signal deconvolution, chemical derivatization or column degradation [100];
- due to indistinct mass spectra as a result of trace analytes or saturated analytes;
- in the case of RI comparison due to subtle differences in stationary phases [12];
- additionally, erroneous and misdirecting entries in mass spectral databases, such as NIST or Fiehn Lib, are possible;
- and especially due to uninformative, identical or highly similar mass spectra e.g. in case of isomeric compounds.

Consequently, in the case of the sugar profiling with many structurally similar isomers and almost identical mass spectra, unambiguous identification was performed using standard compounds (section 3.2-3.3, 4). Even so, both in case of the sugar profiling and VOC analysis several compounds remained unknown. Dunn *et al.* [102] found that using EI-MS based databases only about half of the detected metabolites can be annotated using library matching. So, how can the number of annotated metabolites

be increased? In principal, mass spectral databases have to be expanded in future with more and more metabolites and this not only for EI-MS based databases, but also for new instrumental set-ups like high-resolution/high accuracy mass analyzers, for example Orbitrap or Q-TOF, as well as CI-MS based databases [100]. In general, EI-MS spectra can give hints for elucidation of the structure of an unknown compound (e.g. about functional groups) [2, 12, 100], especially if additional information such as RI or structured retention in the case of GC×GC (section 4) is available. However, this requires time and considerable chemical and analytical expertise regarding fragmentation reactions and fragment structure [12]. Otherwise, high-resolution MS (such as Orbitrap, TOF or Q-TOF) enable the calculation of the empirical chemical formula of a metabolite via accurate mass and relative isotopic abundance. Together with the fragmentation, structural elucidation of unknown compounds is possible. Similarly, MS/MS experiments can be used to gain information on the fragmentation behavior and therefore about the structure. Both, calculation of the empirical chemical formula and also MS/MS experiments necessitate an abundant detection of the molecular ion or pseudo-molecular ion, which is usually of low intensity or not present at all in EI-MS spectra of trimethylsilyl derivatives. Consequently, alternative soft ionization methods resulting in an abundant molecular ion are necessary, such as positive ion chemical ionization CI(+), cold or soft EI (lower ionization energy), field ionization, photo ionization, electrospray ionization or atmospheric pressure chemical ionization (APCI) [101]. Nevertheless, all the above mentioned possibilities will reach their limits in the case of structural isomers such as sugar compounds, because e.g. glucose and galactose will result in the same molecular ion and highly similar fragmentation patterns. In principal, ion mobility spectrometry (IMS) coupled to MS, enabling separation of ionized molecules depending on their mobility in a carrier buffer gas, might be an interesting method for the separation of such isomers [103, 104]. Nonetheless, these problems of structural isomers highlight the necessity of chromatographic separation, which is the reason why the original one-dimensional GC-MS sugar profiling method was transformed towards a two-dimensional GC×GC-MS method, which enabled better separation performance, higher sensitivity and additional information on subclasses due to structured retention (e.g. polyol vs. monosaccharide vs. sugar acid; section 4).

Comparing physiological and mechanistic links revealed by untargeted metabolomics, metabolite profiling approaches and targeted analysis, Sévin *et al.* [7] found that most mechanistic links were revealed by hypothesis-driven targeted analysis or narrower metabolite profiling approaches. This can most likely be attributed to the narrower perspective on a specific pathway or substance class in hypothesis-driven approaches. Therefore, a deeper elucidation and interpretation of metabolic pathways is possible, a typical example is the usage of lipidomics [42]. Often metabolite profiling approaches enable a more coherent interpretation of the relations between the covered metabolite profile and a biological question [7, 41]. Subtle changes in the metabolome due to, for example the diet, are unlikely to be discovered using standard untargeted metabolomics approaches, these often necessitate more specified metabolite profiling approaches [53]. Nonetheless, untargeted metabolomics approaches often reveal unexpected and important insights, but strategies for the interpretation of highly complex metabolic patterns and integration of results into mechanistic physiological pathways still need to be developed and expanded, e.g., by using novel computational approaches [7]. Additionally, untargeted metabolomics is considered as a hypothesis-generating approach and most often results will lead to the development of a targeted method to confirm a hypothesis generated by an untargeted metabolomics approach.

Discovery of markers for dietary intake

The putative markers identified in this work (e.g. mannoheptulose and perseitol for avocado consumption, lactose and galactose for dairy product consumption, 3,4-dimethyl-2,5-furandione for coffee consumption or trehalose for the differentiation of time-courses after an OGTT in healthy, prediabetic and diabetic volunteers; sections 3.1-3.3) were identified in a data-driven or hypothesis-generating approach. While so far most markers used in practice were identified hypothesis-driven, this might change in the future and more and more markers will be identified using hypothesis-generating metabolomics approaches. However, metabolomics approaches can help also for hypothesis-driven identification of potential dietary markers by building databases with metabolite profiles of the foods themselves. Using such databases, marker candidates might be identified by comparing the occurrence of a potentially specific food component with its occurrence in other foods. An example for such a database is the FooDB, which is so far the presumably most comprehensive

resource on food constituents, chemistry and biology, covering more than 25 000 native food compounds (macro- and micronutrients), additives and compounds derived from food processing [105, 106]. A crucial point of such databases is the question, whether the information in these databases are only "collected" using algorithms or whether the information is also checked for reliability, which is a demanding and labor-intensive part. Other relevant databases are national nutrient databases such as the German nutrient database, which is supervised and quality-controlled by the Max Rubner-Institut [107]. Furthermore, more specified databases such as the Phenol-Explorer for polyphenols exist [108]. In the metabolomics workgroup of the MRI, also first steps towards building a metabolite profile database of fruit and vegetables were performed by a comprehensive screening of primary metabolites in overall 33 fruit varieties and 41 vegetable varieties (see list of supervised master and diploma theses). A good example, how such metabolite profile databases might enable the hypothesis-driven identification of dietary markers, is pinitol for soy. During the comprehensive screening of 41 vegetable varieties in our workgroup, it was detected in high amounts in soy beans, but not in other vegetable varieties (data not shown). Within one of the intervention studies of the FoodBAll consortium, Münger *et al.* [70] were able to identify pinitol as a candidate marker for a soy drink. This clearly demonstrates the benefits of expanding the application of metabolite profile databases.

Validation of markers for dietary intake

Validation is a vital step before any marker can be used in practice, e.g., to objectively measure dietary intake or as a clinical marker. As described in the introduction (section 1.2.2), validation serves to ascertain, which confounding factors contribute and how strong their contribution to the variability of the marker is, meaning the consumed amount of a certain food. Dragsted *et al.* [74] proposed the first comprehensive scheme for validation of dietary markers within the frame of the FoodBAll project. The scheme is meant to demonstrate both, the current level of validation and also the necessity for additional studies for full validation of a candidate marker. The validation concept includes eight criteria (section 1.2.2) provided as questions, which can be answered with yes, no or data not available (Supplemental Table S5.1). However, answering these questions unambiguously might be difficult and the particular value of each criterion depends on the intended

purpose of the marker. Therefore, explanatory comments and conditions under which these questions were answered, should be added to substantiate the degree of the validation [74]. The exemplary application of the validation scheme on selected candidate markers identified in this thesis (mannoheptulose and perseitol for avocado consumption, galactose and lactose for dairy product consumption, xylitol and ethyl-β-glucuronide for alcohol consumption and 3,4-dimethyl-2,5-furandione for coffee consumption) is depicted in Figure 5.1. Additionally, in Supplemental Table 5.1 the concrete answering of the validation questions is provided, exemplified for mannoheptulose and perseitol as putative markers for avocado consumption (section 3.2). In the figure, some critical points are marked by half circles, because an unambiguous answer with yes, no or no data available was not possible here.

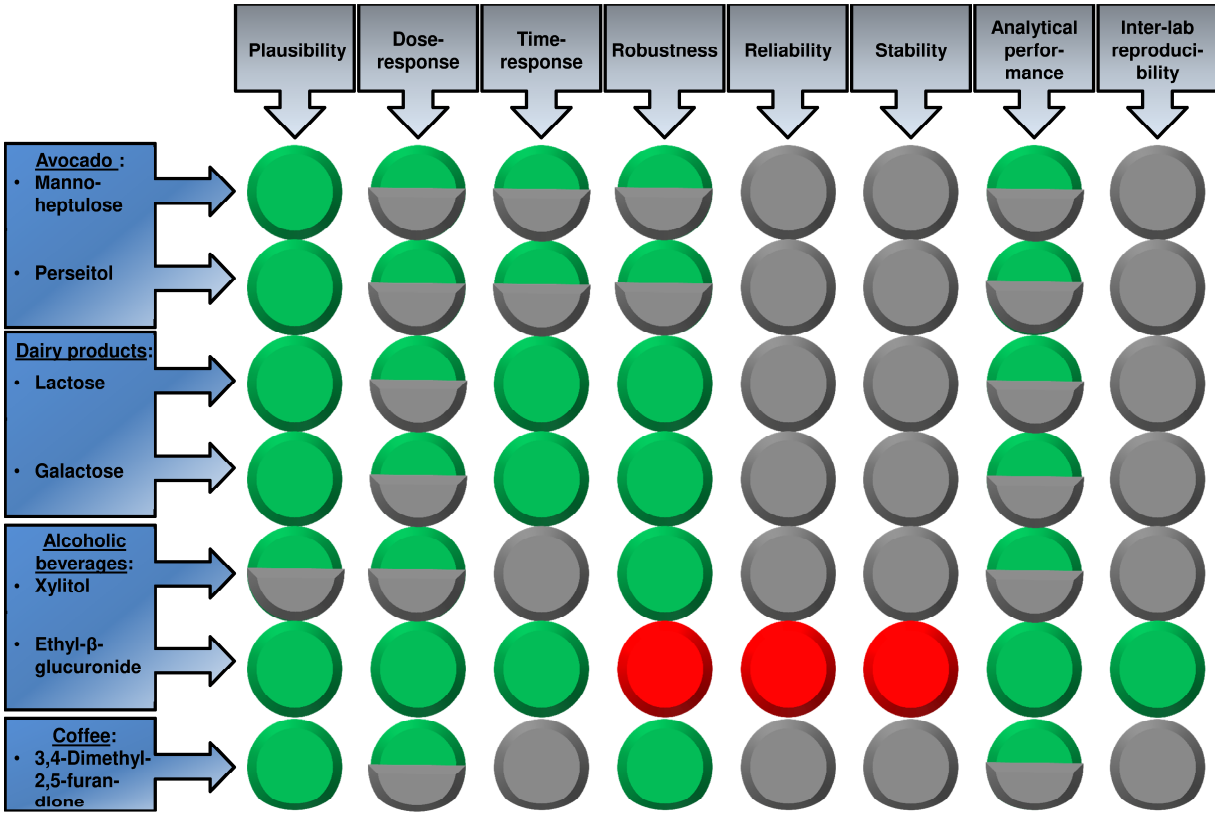


Figure 5.1. Application of the validation scheme developed by Dragsted *et al.* [74] on a selection of putative markers identified in this thesis. The eight validation criteria are depicted in columns and the putative markers in rows. Green circles: validation question answered with yes; red circles: validation question answered with no; grey circles: no data available; half circles: conditional yes, but more data necessary; answer is dependent on the strictness in application of the scheme.

In general, the proposed eight criteria meet the most important aspects to validate candidate markers of dietary intake. Some ambiguities with respect to the application

of the validation scheme are discussed in the following paragraphs together with the general applicability and advantages.

As pointed out above, building and expanding comprehensive metabolite profile databases will help to establish plausibility and specificity of putative markers in the future. It is somewhat surprising that specificity as such, has not been included as a separate criterion in the validation scheme of Dragsted *et al.* [74], but it is included within the criteria plausibility and robustness. This might cause confusion with respect to the definition and application of these criteria. For example, mannitol was associated with mushroom intake in the cross-sectional KarMeN study (300 free-living individuals with a highly diverse diet) and therefore can be considered as robust (section 3.2). Additionally, as a compound naturally occurring in mushrooms, it can also be considered as plausible. However, it cannot be considered as specific, due to its occurrence in a number of other fruit and vegetables (screening of fruit and vegetables, section 3.2), therefore, neither a quantitative nor a qualitative assessment of mushroom consumption is possible. In this respect, it would be more conclusive to classify specificity as a separate criterion to avoid confusion.

With respect to specificity, another interesting aspect is the use of single markers vs. multiple markers for a certain food, which so far is practically unexplored [56]. In the case of multiple markers, specificity, reliability and robustness might be achieved through the combination of compounds. Statistical methods enabling the inclusion of multiple compounds as potential markers for one food should both help to select new panels of markers and also to strengthen their validity through complementary combination. For example, stepwise logistic regression analysis enables the inclusion of more than one metabolite in marker panels and regression against dietary intake data for identification of multiple marker panels; other approaches include multivariate strategies such as PCA, PLS, random forest or support vector machine algorithms [56]. Multiple marker panels will enable more accurate estimation of dietary intake, especially if additionally combined with data from traditional dietary assessment methods [56]. Within this thesis, multiple markers for one food were also detected, e.g., mannopheptulose and perseitol for avocado consumption. Mannoheptulose is less specific than perseitol due to its occurrence in other foods. However, it is more abundant in urine and thus detected more frequently. Therefore, the complementary combination of both avocado markers might result in a more

accurate quantification, where the lower specificity of mannoheptulose is attenuated by additional consideration of perseitol. However, statistical approaches in this direction still need to be developed.

A criterion causing ambiguities in its interpretation and application is reliability. Dragsted *et al.* [74] require confirmation studies with highly controlled settings (e.g. supervised intake with known exact amounts) to compare the performance of the putative marker to indicate true dietary intake against a gold standard. However, in practice, for almost no newly discovered dietary marker, a gold standard or reference method for comparison is available. Thus, the current best practice is comparison with traditional dietary assessment methods or, if several putative markers for the same food were discovered at the same time, they may be validated for reliability against each other [74]. However, Dragsted *et al.* urge for care in interpreting reliability in such cases, as both approaches are not without pitfalls, e.g. misreporting in case of comparison against traditional dietary assessment methods (see introduction section 1.2.1 and 1.2.2). Consequently, the question on how to exactly apply the criterion reliability remains unclear.

In general, more rigid criteria to unambiguously answer the validation questions with yes or no for each criterion might be helpful to decide, if a criterion can be considered as fulfilled or not (Figure 5.1). For example, in the case of dose-response relationship, it should be differentiated between the establishment of the criterion in a cross-sectional study or in an intervention study, especially in the light of the quantitative assessment of dietary intake. In a cross-sectional study, bias through the background diet, the utilization of self-reported intake and the necessity to define food variables (e.g., not only black coffee, but also milk coffee, cappuccino etc. were included for coffee consumption; section 3.1) influence quantification of dietary intake. In contrast, the study food during an intervention is well-defined and different intake amounts are compared with a normally not interfering background diet. Therefore, a classification within each criterion, together with more rigid criteria is necessary to determine validity. Such criteria were already mentioned by Dragsted *et al.*; however, further work will be necessary to define such a classification, depending on the different intended purposes of markers [74]. Another aspect that should be considered to define whether a criterion can be classified as fulfilled, is the quality and the significance of the studies used for validation. This includes aspects such as

e.g., number of study participants, well-defined intervention food, standardization of the background diet, run-in and wash-out periods during a study, etc. For example, mannoheptulose and perseitol as putative markers for avocado consumption were identified in the cross-sectional KarMeN study and therefore, can be considered as robust; however, only nine out of 300 participants consumed avocado. Also, with respect to the criteria time- and dose-response, the intervention study with avocado by Wamelink *et al.* [109] was performed with only three volunteers. Considering the low number of volunteers in both studies, the need for studies with more volunteers should be highlighted here (see half circles Figure 5.1).

Another ambiguity concerns the last three validation criteria addressing the validity of the analytical method and the clear differentiation from the first five validation criteria addressing biological validity. Within their validation scheme, Dragsted *et al.* [74] address both biological and analytical validity; however, both should be viewed as independent aspects. On the one side, the first five criteria and in part also the criterion stability (chemical and biological) might be considered as intrinsic criteria of a putative marker, addressing whether a metabolite fulfills biological validity. On the other side, the analytical performance addresses requirements, which are necessary to obtain valid results while discovering new markers and also while applying a biologically valid marker. Therefore, analytical performances (i.e. analytical validity) is a requirement for both, the method used for marker discovery (likely untargeted), and the method used for application of a biologically valid marker (likely targeted). However, Dragsted *et al.* [74] consider analytical performance mainly related to the question, whether qualitative or quantitative analysis is possible using a putative marker or whether the development of a quantitative targeted method is feasible. Yet, this remains unclear, if only the validation questions (see Supplemental Table 5.1) are considered. Despite the fact that analytical validity should be viewed as an independent step, some aspects of the criteria for biological validity are also dependent on the actually used analytical method. For example, defining the range of intake that could be measured with a certain putative marker does not only depend on the marker itself, but also on the limit of detection and linearity range of the applied analytical method. Overall, the application and interpretation of the criterion analytical performance seems incomplete and unclear. Combining different methods in the literature to validate a putative marker has to be reviewed and interpreted

critically, because a full validation, both biologically and analytically, should be carried out with the method intended for use in practice. Overall, all aspects causing ambiguities (as discussed in the paragraphs above) should be addressed in a further development of the validation scheme.

After discussion of the general applicability of the scheme [74], the practical application of the validation scheme on the putative marker candidates identified in this thesis, are discussed in the following paragraph. The scheme clearly revealed the necessity for further studies to validate the candidate markers discovered in this thesis (Figure 5.1 and Supplemental Table 5.1). So far, all marker candidates might only be considered as plausible qualitative markers of dietary intake, with the exception of ethyl-β-glucuronide, which is routinely and quantitatively used for the proof of abstinence [110]. However, its application is questionable due to problems with respect to robustness, reliability and stability (section 3.2). For evaluation of the putative markers identified in this thesis as quantitative markers of dietary intake, further controlled intervention studies with different doses of food intake and no influencing background diet will be necessary. So far, dose-response was established in a cross-sectional study, thus only half circles were applied in Figure 5.1. With the exception of 3,4-dimethyl-2,5-furandione and xylitol, time-response after a single meal was established for the candidate markers in the literature [70, 109, 111] (also sections 3.1-3.2), although in the case of mannoheptulose and perseitol only a limited number of participants was available (Figure 5.1). No literature can be found about the influence of repeated intake on the putative markers (e.g. inhibition or activation of enzymes, accumulation etc.), therefore highlighting the need for further investigations in this respect. Robustness can be assumed to a certain degree for all putative markers, due to the establishment within the cross-sectional KarMeN study with unrestricted background diet, except for mannoheptulose and perseitol (limited number of participants) and ethyl-β-glucuronide (section 3.2). Additionally, reliability (i.e. comparison of the marker performance with other described markers or a dietary assessment method) has to be evaluated in the future. For example, in the case of coffee, reliability might be assessed via measuring caffeine metabolites or trigonelline [68] and in the case of milk via measuring pentadecanoic acid in serum and adipose tissue for total dairy fat intake [112]. Considering analytical validity (stability, analytical performance and inter-

laboratory reproducibility), first targeted methods for the application of the candidate markers have to be developed for all candidate markers.

Overall, it can be concluded that in principle both metabolite profiling approaches (sugar profiling and VOC analysis) applied in this thesis enabled the detection of putative markers for dietary intake, which after further validation might be applied for a more objective measurement of dietary intake, monitoring of compliance in studies or validation of traditional dietary assessment methods. Using these metabolite profiling approaches, further potential dietary markers will be discovered in the future (section 3.1 Table 4: e.g. 2-pentylfurane for bread, section 3.2 Table 4 and Figure 4, or see section 4 Figure 3.9: putative markers for apple and coke consumption). Expanding the number of available validated dietary markers is imperative to improve the understanding of the influence of the diet on health and disease. For example, measurement errors in self-reported intake of sugar (due to misreporting) prevent finding consistent epidemiologic evidence on the association between sugar and disease risk (e.g. for cardiovascular disease (CVD), type 2 diabetes and cancer links have been ambiguous and inconclusive) [58]. Here, dietary markers for sugar intake might enable obtaining objective and accurate estimates of intake and circumvent problems associated e.g. with underreporting [113, 114].

Another point relevant for obscured and inconsistent associations between diet and health is the complexity of the diet itself. Any food consists of several hundred compounds, and each of them might have positive, negative, synergistic or cumulative health effects that may depend on the food matrix composition. Therefore, it is highly recommended to analyze the composition of any food under investigation for a health effect. Metabolomics approaches (untargeted metabolomics or metabolite profiling approaches) can help to facilitate the understanding of the complex and dynamic responses of the human metabolism to nutrition in two complementary ways. First, metabolite profile databases of the food itself will help understanding the complexity of the food itself (see above).

Discovery of markers for health status

Second, metabolomics approaches are of high interest to identify new biomarkers for defining or describing health status and early disease states. To demonstrate this, the developed semitargeted sugar profiling method was applied to investigate the sugar profile in plasma of healthy, prediabetic and diabetic volunteers after an OGTT

(section 3.3). Interestingly, although type 2 diabetes is primarily characterized by a disturbance in glucose metabolism, other sugar compounds have only rarely been investigated. The application of the sugar profiling method enabled the detection of a complex plasma sugar profile and revealed further sugar compounds next to glucose showing characteristic differences depending on health status (section 3.3).

One could argue whether new biomarkers for diagnosis of T2DM are required, given that well-established and easily measurable biomarkers such as glucose and HbA_{1C} already exist. However, metabolomics approaches enable generating whole signatures of multiple metabolites, and therefore have the potential for sensitive and accurate detection of a disease state [7]. This is especially relevant for early diagnosis of disease, where additional parameters might result in a better prediction of the risk for certain diseases. Early diagnosis of T2DM is highly relevant, because during this period effective preventive measures are available; remarkably enough, still half of the people having T2DM remain undiagnosed for years [86] and during this time many of the later emerging complications already develop. The prediction of the development and severity of these complications might be further facilitated by metabolomics approaches. In multifactorial diseases such as T2DM, the phenotype is highly dynamic and complex because various factors (genetic, life style, environmental, etc.) have an influence on manifestation and lead to several different subtypes [3]. The classification of metabolic subtypes of a disease via multiple metabolite signatures might open up the possibility of personalized treatment.

Detecting metabolic subtypes in a dynamic system is best evaluated during a metabolic challenge, such as the OGTT performed for diagnosis of diabetes [3]. Performing the OGTT with healthy, prediabetic and diabetic subjects and measuring the resulting response in the sugar profile showed that, contrary to expectations, a number of further sugar compounds showed characteristic health status-dependent changes (section 3.3), which in future may serve as biomarkers of early stage T2DM. Additionally, using the sugar profile might enable the differentiation of subtypes in the disease or the likelihood for certain complications, enabling better prevention, monitoring and therapy. To investigate this, an intervention study with a large collective of participants (prediabetic and diabetic) performing an OGTT would be necessary. In our study, no assessment of subtypes was possible due to the limited number of volunteers (only 15 healthy, 15 prediabetic and 11 diabetic volunteers).

And similarly to dietary markers, an extensive evaluation and validation of the changes in the sugar profile has to be performed, before the sugar profile as a whole, or single sugar compounds might be used as biomarkers. The first and most important step would be the investigation of the causal link/plausibility between the characteristic differences in health status observed, for example for mannose, trehalose or xylose (section 3.3) and therefore, the understanding of how they might serve as biomarkers. Additional research/interventions with stable-isotope labeled glucose will help to verify, whether these compounds were derived from glucose during the OGTT and give first hints on the mechanistic links of these sugar compounds and T2DM disease state. Further validation criteria address similar aspects as those for dietary markers [53]: specificity for the disease, defining threshold levels for disease stage, defining time-lines/levels/thresholds for progression of the disease, robustness in the population, reliability of diagnosing a disease and analytical aspects (stability of the marker compounds, analytical performance and inter-laboratory reproducibility of the final biomarker-assay) need to be considered.

Overall, the sugar profiling method enabled unexpected insights into human physiology, even though no explanation of the underlying mechanisms can be provided yet. Maltose enabled the differentiation of the menopausal state in women (section 3.2), but the link with vaginal glycogen and Lactobacilli counts in vaginal fluid as a result of reduced estrogen concentrations needs to be confirmed in future. The occurrence of psicose in urine and also the increase of psicose after an OGTT, the causality underlying the excretion of xylitol after alcohol consumption, and also the above mentioned sugar compounds potentially indicating diabetic state, both in fasting state and after an OGTT (e.g. trehalose; section 3.3), raise a number of questions with respect to the depth of our understanding of carbohydrate metabolism in health and disease.

In conclusion, both metabolite profiling approaches developed in this thesis, the semitargeted sugar profiling and the VOC analysis, are valuable and useful tools allowing the identification of markers of dietary intake and also markers of health status and disease states as demonstrated in application studies.

6 REFERENCES

1. Fiehn O. Metabolomics — the link between genotypes and phenotypes. Edtion ed. In: Town C, ed. Functional Genomics. Dordrecht: Springer Netherlands, 2002:155-71.
2. Dettmer K, Aronov PA, Hammock BD. Mass spectrometry-based metabolomics. Mass spectrom Rev 2007;26(1):51-78.
3. Ramautar R, Berger R, van der Greef J, Hankemeier T. Human metabolomics: strategies to understand biology. Curr Opin Chem Biol 2013;17(5):841-6.
4. Koek MM, Jellema RH, van der Greef J, Tas AC, Hankemeier T. Quantitative metabolomics based on gas chromatography mass spectrometry: status and perspectives. Metabolomics 2011;7(3):307-28.
5. Zhang A, Sun H, Wang P, Han Y, Wang X. Recent and potential developments of biofluid analyses in metabolomics. J Proteomics 2012;75(4):1079-88.
6. Krastanov A. Metabolomics—The State of Art. Biotechnol Biotechnol Equip 2010;24(1):1537-43.
7. Sévin DC, Kuehne A, Zamboni N, Sauer U. Biological insights through nontargeted metabolomics. Curr Opin Biotechnol 2015;34:1-8.
8. Hollywood K, Brison DR, Goodacre R. Metabolomics: Current technologies and future trends. Proteomics 2006;6(17):4716-23.
9. Wishart DS, Feunang YD, Marcu A, Guo AC, Liang K, Vázquez-Fresno R, Sajed T, Johnson D, Li C, Karu N, et al. HMDB 4.0: the human metabolome database for 2018. Nucleic Acids Res 2018;46(Database issue):D608-D17.
10. Brown M, Dunn WB, Ellis DI, Goodacre R, Handl J, Knowles JD, O'Hagan S, Spasić I, Kell DBJM. A metabolome pipeline: from concept to data to knowledge. Metabolomics 2005;1(1):39-51.
11. Ryan D, Robards K. Metabolomics: The Greatest Omics of Them All? Anal Chem 2006;78(23):7954-8.
12. Ulaszewska MM, Weinert CH, Trimigno A, Portmann R, Andres Lacueva C, Badertscher R, Brennan L, Brunius C, Bub A, Capozzi F, et al. Nutrimetabolomics: An Integrative Action for Metabolomic Analyses in Human Nutritional Studies. Mol Nutr Food Res 2019;63:1800384.
13. Gertsman I, Barshop BA. Promises and pitfalls of untargeted metabolomics. J Inherited Metab Dis 2018;41(3):355-66.
14. Dunn WB, Bailey NJC, Johnson HE. Measuring the metabolome: current analytical technologies. Analyst 2005;130(5):606-25.
15. Alonso A, Marsal S, Julià A. Analytical Methods in Untargeted Metabolomics: State of the Art in 2015. Front Bioeng Biotechnol 2015;3(23).
16. Mastrangelo A, Ferrarini A, Rey-Stolle F, García A, Barbas C. From sample treatment to biomarker discovery: A tutorial for untargeted metabolomics based on GC-(EI)-Q-MS. Anal Chim Acta 2015;900:21-35.
17. Wachsmuth CJ, Vogl FC, Oefner PJ, Dettmer K. CHAPTER 5 Gas Chromatographic Techniques in Metabolomics. Edtion ed. Chromatographic Methods in Metabolomics: The Royal Society of Chemistry, 2013:87-113.

18. Mills GA, Walker V. Headspace solid-phase microextraction procedures for gas chromatographic analysis of biological fluids and materials. J Chromatogr A 2000;902(1):267-87.
19. Soria AC, García-Sarrió MJ, Sanz ML. Volatile sampling by headspace techniques. TrAC Trends Anal Chem 2015;71:85-99.
20. Arthur CL, Pawliszyn J. Solid phase microextraction with thermal desorption using fused silica optical fibers. Anal Chem 1990;62(19):2145-8.
21. Bojko B, Reyes-Garcés N, Bessonneau V, Goryński K, Mousavi F, Souza Silva EA, Pawliszyn J. Solid-phase microextraction in metabolomics. TrAC, Trends Anal Chem 2014;61:168-80.
22. Cozzolino R, De Magistris L, Saggese P, Stocchero M, Martignetti A, Di Stasio M, Malorni A, Marotta R, Boscaino F, Malorni L. Use of solid-phase microextraction coupled to gas chromatography–mass spectrometry for determination of urinary volatile organic compounds in autistic children compared with healthy controls. Anal Bioanal Chem 2014;406(19):4649-62.
23. Monteiro M, Carvalho M, Henrique R, Jerónimo C, Moreira N, de Lourdes Bastos M, de Pinho PG. Analysis of volatile human urinary metabolome by solid-phase microextraction in combination with gas chromatography–mass spectrometry for biomarker discovery: Application in a pilot study to discriminate patients with renal cell carcinoma. Eur J Cancer 2014;50(11):1993-2002.
24. Want EJ, Nordström A, Morita H, Siuzdak G. From Exogenous to Endogenous: The Inevitable Imprint of Mass Spectrometry in Metabolomics. J Proteome Res 2007;6(2):459-68.
25. Giddings J. Concepts and comparisons in multidimensional separation. J High Resolut Chromatogr 1987;10(5):319-23.
26. Weinert C. Über die Eignung der umfassenden zweidimensionalen Gaschromatografie mit Quadrupol-MS-Detektion (GC×GC-qMS) für die Durchführung von Metabolomanalysen humaner und pflanzlicher Matrizes. Dissertation am KIT 2014.
27. Ryan D, Morrison P, Marriott P. Orthogonality considerations in comprehensive two-dimensional gas chromatography. J Chromatogr A 2005;1071(1–2):47-53.
28. Mondello L, Tranchida PQ, Dugo G, Dugo P. Comprehensive Two-Dimensional Gas Chromatography-Mass Spectrometry: A Review. Mass Spectrom Rev 2008;27(101-124).
29. Adahchour M, Beens J, Vreuls RJJ, Brinkman UAT. Recent developments in comprehensive two-dimensional gas chromatography (GC×GC) I. Indtroduction and instrumental set-up. Trends Anal Chem 2006;25(5):438-54.
30. Dallüge J, Vreuls RJJ, Beens J, Brinkman UAT. Optimization and characterization of comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometric detection (GC×GC–TOF MS). J Sep Sci 2002;25(4):201-14.
31. Weinert CH, Egert B, Kulling SE. On the applicability of comprehensive two-dimensional gas chromatography combined with a fast-scanning quadrupole mass spectrometer for untargeted large-scale metabolomics. J Chromatogr A 2015;1405:156-67.

32. Adahchour M, Beens J, Vreuls RJJ, Brinkman UAT. Recent developments in comprehensive two-dimensional gas chromatography (GCxGC) II. Modulation and detection. Trends Anal Chem 2006;25(6):540-53.
33. Adahchour M, M B, HU B, Vreuls RJJ, AM B, Brinkman UAT. Comprehensive two-dimensional gas chromatography coupled to a rapid-scanning quadrupol mass spectrometer: principles and applications. J Chromatogr A 2005;1067:245-54.
34. Dallüge J, Beens J, Brinkman UAT. Comprehensive two-dimensional gas chromatography: a powerful and versatile analytical tool. J Chromatogr A 2003;1000:69-108.
35. Zivkovic AM, German JB. Metabolomics for assessment of nutritional status. Curr Opin Clin Nutr Metab Care 2009;12(5):501-7.
36. German JB, Bauman DE, Burrin DG, Failla ML, Freake HC, King JC, Klein S, Milner JA, Pelto GH, Rasmussen KM, et al. Metabolomics in the Opening Decade of the 21st Century: Building the Roads to Individualized Health. J Nutr 2004;134(10):2729-32.
37. Zeisel SH, Freake HC, Bauman DE, Bier DM, Burrin DG, German JB, Klein S, Marquis GS, Milner JA, Pelto GH, et al. The Nutritional Phenotype in the Age of Metabolomics. J Nutr 2005;135(7):1613-6.
38. Bingham SA. Biomarkers in nutritional epidemiology. Public Health Nutr 2002;5(6a):821-7.
39. Favé G, Beckmann ME, Draper JH, Mathers JC. Measurement of dietary exposure: a challenging problem which may be overcome thanks to metabolomics? Genes Nutr 2009;4(2):135-41.
40. Wishart DS. Metabolomics: applications to food science and nutrition research. Trends Food Sci Technol 2008;19(9):482-93.
41. Scalbert A, Brennan L, Fiehn O, Hankemeier T, Kristal B, van Ommen B, Pujos-Guillot E, Verheij E, Wishart D, Wopereis S. Mass-spectrometry-based metabolomics: limitations and recommendations for future progress with particular focus on nutrition research. Metabolomics 2009;5(4):435-58.
42. Gibney MJ, Walsh M, Brennan L, Roche HM, German B, van Ommen B. Metabolomics in human nutrition: opportunities and challenges. Am J Clin Nutr 2005;82(3):497-503.
43. Subar AF, Kipnis V, Troiano RP, Midthune D, Schoeller DA, Bingham S, Sharbaugh CO, Trabulsi J, Runswick S, Ballard-Barbash R, et al. Using Intake Biomarkers to Evaluate the Extent of Dietary Misreporting in a Large Sample of Adults: The OPEN Study. Am J Epidemiol 2003;158(1):1-13.
44. Kuhnle GGC. Nutritional biomarkers for objective dietary assessment. J Sci Food Agric 2012;92(6):1145-9.
45. Manach C, Hubert J, Llorach R, Scalbert A. The complex links between dietary phytochemicals and human health deciphered by metabolomics. Mol Nutr Food Res 2009;53(10):1303-15.
46. Potischman N, Freudenheim JL. Biomarkers of Nutritional Exposure and Nutritional Status: An Overview. J Nutr 2003;133(3):873S-4S.
47. Koulman A, Volmer DA. Perspectives for metabolomics in human nutrition: an overview. BNF Nutr Bull 2008;33(4):324-30.

48. Scalbert A, Brennan L, Manach C, Andres-Lacueva C, Dragsted LO, Draper J, Rappaport SM, van der Hooft JJ, Wishart DS. The food metabolome: a window over dietary exposure. Am J Clin Nutr 2014.
49. Watkins SM, German JB. Toward the implementation of metabolomic assessments of human health and nutrition. Curr Opin Biotechnol 2002;13(5):512-6.
50. Gao Q, Pratico G, Scalbert A, Vergeres G, Kolehmainen M, Manach C, Brennan L, Afman LA, Wishart DS, Andres-Lacueva C, et al. A scheme for a flexible classification of dietary and health biomarkers. Genes Nutr 2017;12:34.
51. Corella D, Ordovás JM. Biomarkers: background, classification and guidlines for applications in nutritional epidemiology. Nutr Hosp 2015;31:177-88.
52. Dragsted LO, Gao Q, Pratico G, Manach C, Wishart DS, Scalbert A, Feskens EJM. Dietary and health biomarkers-time for an update. Genes Nutr 2017;12:7.
53. de Vries J, Antoine JM, Burzykowski T, Chiodini A, Gibney M, Kuhnle G, Meheust A, Pijls L, Rowland I. Markers for nutrition studies: review of criteria for the evaluation of markers. Eur J Nutr 2013;52(7):1685-99.
54. WHO. Evironmental Health Criteria 237. Principles for evaluating health risks in children associated with exposure to chemicals. WHO Press, Geneva, Switzerland, 2006.
55. WHO Internation Programme on Chemical Safety. Biomarkers and Risk Assessment: concepts and Principles. WHO Press, Geneva, Switzerland, 1993.
56. Garcia-Aloy M, Rabassa M, Casas-Agustench P, Hidalgo-Liberona N, Llorach R, Andres-Lacueva C. Novel strategies for improving dietary exposure assessment: Multiple-data fusion is a more accurate measure than the traditional single-biomarker approach. Trends Food Sci Technol 2017;69:220-9.
57. Luceri C, Caderni G, Lodovici M, Spagnesi MT, Monserrat C, Lancioni L, Dolara P. Urinary excretion of sucrose and fructose as a predictor of sucrose intake in dietary intervention studies. Cancer Epidemiol, Biomarkers Prev 1996;5(3):167-71.
58. Tasevska N. Urinary Sugars - A Biomarker of Total Sugars Intake. Nutrients 2015;7(7):5816-33.
59. Joosen AMCP, Kuhnle GGC, Runswick SA, Bingham SA. Urinary sucrose and fructose as biomarkers of sugar consumption: comparison of normal weight and obese volunteers. Int J Obes 2008;32(11):1736-40.
60. Johner SA, Libuda L, Shi L, Retzlaff A, Joslowski G, Remer T. Urinary fructose: a potential biomarker for dietary fructose intake in children. Eur J Clin Nutr 2010;64(11):1365-70.
61. Dresen S, Weinmann W, Wurst FM. Forensic confirmatory analysis of ethyl sulfate—a new marker for alcohol consumption—by liquid-chromatography/electrospray ionization/tandem mass spectrometry. J Am Soc Mass Spectrom 2004;15(11):1644-8.
62. Wurst FM, Kempter C, Metzger J, Seidl S, Alt A. Ethyl glucuronide: a marker of recent alcohol consumption with clinical and forensic implications. Alcohol 2000;20(2):111-6.

63. Wurst FM, Kempter C, Seidl S, Alt A. Ethyl glucuronide--a marker of alcohol consumption and a relapse marker with clinical and forensic implications. Alcohol Alcohol 1999;34(1):71-7.
64. Ross AB. Present status and perspectives on the use of alkylresorcinols as biomarkers of wholegrain wheat and rye intake. J Nutr Metab 2012;2012:462967-.
65. Gibbons H, Michielsen CJR, Rundle M, Frost G, McNulty BA, Nugent AP, Walton J, Flynn A, Gibney MJ, Brennan L. Demonstration of the utility of biomarkers for dietary intake assessment; proline betaine as an example. Mol Nutr Food Res 2017;61(10):1700037.
66. O'Sullivan A, Gibney MJ, Brennan L. Dietary intake patterns are reflected in metabolomic profiles: potential role in dietary assessment studies. Am J Clin Nutr 2011;93(2):314-21.
67. Andersen M-BS, Kristensen M, Manach C, Pujos-Guillot E, Poulsen SK, Larsen TM, Astrup A, Dragsted LJA, Chemistry B. Discovery and validation of urinary exposure markers for different plant foods by untargeted metabolomics. Anal Bioanal Chem 2014;406(7):1829-44.
68. Rothwell JA, Madrid-Gambin F, Garcia-Aloy M, Andres-Lacueva C, Logue C, Gallagher AM, Mack C, Kulling SE, Gao Q, Praticò G, et al. Biomarkers of intake for coffee, tea, and sweetened beverages. Genes Nutr 2018;13(1):15.
69. Praticò G, Gao Q, Manach C, Dragsted LOJG, Nutrition. Biomarkers of food intake for Allium vegetables. Genes Nur 2018;13(1):34.
70. Münger LH, Trimigno A, Picone G, Freiburghaus C, Pimentel G, Burton KJ, Pralong FP, Vionnet N, Capozzi F, Badertscher R, et al. Identification of Urinary Food Intake Biomarkers for Milk, Cheese, and Soy-Based Drink by Untargeted GC-MS and NMR in Healthy Humans. J Proteome Res 2017.
71. Mack CI, Weinert CH, Egert B, Ferrario PG, Bub A, Hoffmann I, Watzl B, Daniel H, Kulling SE. The complex human urinary sugar profile: determinants revealed in the cross-sectional KarMeN study. Am J Clin Nutr 2018;108(3):502-16.
72. O'Gorman A, Brennan L. The role of metabolomics in determination of new dietary biomarkers. Proc Nutr Soc 2017;76(3):295-302.
73. Kroke A, Klipstein-Grobusch K, Hoffmann K, Terbeck I, Boeing H, Helander A. Comparison of self-reported alcohol intake with the urinary excretion of 5-hydroxytryptophol:5-hydroxyindole-3-acetic acid, a biomarker of recent alcohol intake. Br J Nutr 2001;85(5):621-7.
74. Dragsted LO, Gao Q, Scalbert A, Vergeres G, Kolehmainen M, Manach C, Brennan L, Afman LA, Wishart DS, Andres Lacueva C, et al. Validation of biomarkers of food intake-critical assessment of candidate biomarkers. Genes Nutr 2018;13:14.
75. Newgard CB. Metabolomics and Metabolic Diseases: Where Do We Stand? Cell Metab 2017;25(1):43-56.
76. Zhang A, Sun H, Yan G, Wang P, Wang X. Metabolomics for Biomarker Discovery: Moving to the Clinic BioMed Res Internat 2015;2015:6.
77. Gowda GAN, Zhang S, Gu H, Asiago V, Shanaiah N, Raftery D. Metabolomics-based methods for early disease diagnostics. Expert Rev Mol Diagn 2008;8(5):617-33.

78. Suárez M, Caimari A, del Bas JM, Arola L. Metabolomics: An emerging tool to evaluate the impact of nutritional and physiological challenges. TrAC Trends Anal Chem 2017;96:79-88.
79. van Ommen B, Keijer J, Heil SG, Kaput J. Challenging homeostasis to define biomarkers for nutrition related health. Mol Nutr Food Res 2009;53(7):795-804.
80. van Ommen B, van der Greef J, Ordovas JM, Daniel HJG, Nutrition. Phenotypic flexibility as key factor in the human nutrition and health relationship. Genes Nutr 2014;9(5):423.
81. Krug S, Kastenmuller G, Stuckler F, Rist MJ, Skurk T, Sailer M, Raffler J, Romisch-Margl W, Adamski J, Prehn C, et al. The dynamic range of the human metabolome revealed by challenges. FASEB journal 2012;26(6):2607-19.
82. Chen J, Zhao X, Fritsche J, Yin P, Schmitt-Kopplin P, Wang W, Lu X, Häring HU, Schleicher ED, Lehmann R, et al. Practical Approach for the Identification and Isomer Elucidation of Biomarkers Detected in a Metabonomic Study for the Discovery of Individuals at Risk for Diabetes by Integrating the Chromatographic and Mass Spectrometric Information. Anal Chem 2008;80(4):1280-9.
83. Friedrich N. Metabolomics in diabetes research. J Endocrinol 2012;215(1):29-42.
84. Wang-Sattler R, Yu Z, Herder C, Messias AC, Floegel A, He Y, Heim K, Campillos M, Holzapfel C, Thorand B, et al. Novel biomarkers for pre-diabetes identified by metabolomics. Molecular systems biology 2012;8:615.
85. Wörmann K, Walker A, Moritz F, Forcisi S, Tziotis D, Lucio M, Heinzmann SS, Adamski J, Lehmann R, Häring H-U, et al. Revolution in der Diabetesdiagnostik dank -omics – Biomarker mittels Metabolomics. Diabetes aktuell 2012;10(03):129-33.
86. Cho NH, Shaw JE, Karuranga S, Huang Y, da Rocha Fernandes JD, Ohlrogge AW, Malanda B. IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045. Diabetes Res Clin Pract 2018;138:271-81.
87. WHO. WHO Press, Geneva, Switzerland, 2006.
88. Belhassen E, Bressanello D, Merle P, Raynaud E, Bicchi C, Chaintreau A, Cordero C. Routine quantification of 54 allergens in fragrances using comprehensive two-dimensional gas chromatography-quadrupole mass spectrometry with dual parallel secondary columns. Part I: Method development. Flavour Fragr J 2018;33(1):63-74.
89. Debonneville C, Chaintreau A. Quantitation of suspected allergens in fragrances: Part II. Evaluation of comprehensive gas chromatography–conventional mass spectrometry. J Chromatogr A 2004;1027(1):109-15.
90. Taylor CM, Raines JM, Rodriguez JMJFAM. Determination of Polycyclic Aromatic Hydrocarbons in Seafood. Food Anal Methods 2013;6(5):1330-6.
91. Egert B, Weinert CH, Kulling SE. A peaklet-based generic strategy for the untargeted analysis of comprehensive two-dimensional gas chromatography mass spectrometry data sets. J Chromatogr A 2015;1405:168-77.
92. Górecki T, Mostafa A. Sensitivity of Comprehensive Two-dimensional Gas Chromatography (GCXGC) Versus One-dimensional Gas Chromatography (1D GC). LCGC Europe 2013;26(12):672-9.

93. Liu Z, Phillips JB. Sensitivity and detection limit enhancement of gas chromatographic detection by thermal modulation. J Microcol Sep 1994;6(3):229-35.
94. Kinghorn RM, Marriott PJ. Enhancement of Signal-to-Noise Ratios in Capillary Gas Chromatography by Using a Longitudinally Modulated Cryogenic System. J Sep Sci 1998;21(1):32-8.
95. Lei Z, Huhman D, Sumner LW. Mass Spectrometry Strategies in Metabolomics. J Biol Chem 2011;286(29):25435–42.
96. Wu L, Mashego MR, van Dam JC, Proell AM, Vinke JL, Ras C, van Winden WA, van Gulik WM, Heijnen JJ. Quantitative analysis of the microbial metabolome by isotope dilution mass spectrometry using uniformly 13C-labeled cell extracts as internal standards. Anal Biochem 2005;336(2):164-71.
97. Birkemeyer C, Luedemann A, Wagner C, Erban A, Kopka J. Metabolome analysis: the potential of in vivo labeling with stable isotopes for metabolite profiling. Trends Biotechnol 2005;23(1):28-33.
98. Guo K, Li L. Differential 12C-/13C-Isotope Dansylation Labeling and Fast Liquid Chromatography/Mass Spectrometry for Absolute and Relative Quantification of the Metabolome. Anal Chem 2009;81(10):3919-32.
99. Fiehn O, Robertson D, Griffin J, van der Werf M, Nikolau B, Morrison N, Sumner L, Goodacre R, Hardy N, Taylor C, et al. The metabolomics standards initiative (MSI). Metabolomics 2007;3(3):175-8.
100. Vinaixa M, Schymanski EL, Neumann S, Navarro M, Salek RM, Yanes O. Mass spectral databases for LC/MS- and GC/MS-based metabolomics: State of the field and future prospects. TrAC Trends Anal Chem 2016;78:23-35.
101. Strehmel N, Kopka J, Scheel D, Böttcher CJM. Annotating unknown components from GC/EI-MS-based metabolite profiling experiments using GC/APCI(+)-QTOFMS. Metabolomics 2014;10(2):324-36.
102. Dunn WB, Erban A, Weber RJM, Creek DJ, Brown M, Breitling R, Hankemeier T, Goodacre R, Neumann S, Kopka J, et al. Mass appeal: metabolite identification in mass spectrometry-focused untargeted metabolomics. Metabolomics 2013;9(1):44-66.
103. Zhu M, Bendiak B, Clowers B, Hill HH. Ion mobility-mass spectrometry analysis of isomeric carbohydrate precursor ions. Anal Bioanal Chem 2009;394(7):1853-67.
104. Dwivedi P, Bendiak B, Clowers B, Hill HH. Rapid resolution of carbohydrate isomers by electrospray ionization ambient pressure ion mobility spectrometry-time-of-flight mass spectrometry (ESI-APIMS-TOFMS). J Am Soc Mass Spectrom 2007;18(7):1163-75.
105. Wishart D. Systems Biology Resources Arising from the Human Metabolome Project. Edtion ed. In: Suhre K, ed. Genetics Meets Metabolomics: from Experiment to Systems Biology. New York, NY: Springer New York, 2012:157-75.
106. Brouwer-Brolsma EM, Brennan L, Drevon CA, van Kranen H, Manach C, Dragsted LO, Roche HM, Andres-Lacueva C, Bakker SJL, Bouwman J, et al. Combining traditional dietary assessment methods with novel metabolomics techniques: present efforts by the Food Biomarker Alliance. Proc Nutr Soc 2017;76(4):619-27.

107. Hartmann BM, Heuer T, Hoffmann I. The German Nutrient Database: Effect of different versions on the calculated energy and nutrient intake of the German population. J Food Compos Anal 2015;42:26-9.
108. Neveu V, Perez-Jiménez J, Vos F, Crespy V, du Chaffaut L, Mennen L, Knox C, Eisner R, Cruz J, Wishart D, et al. Phenol-Explorer: an online comprehensive database on polyphenol contents in foods. Database 2010;2010:bap024-bap.
109. Wamelink MM, Smith DE, Jansen EE, Verhoeven NM, Struys EA, Jakobs C. Detection of transaldolase deficiency by quantification of novel seven-carbon chain carbohydrate biomarkers in urine. J Inherit Metab Dis 2007;30(5):735-42.
110. Thierauf A, Wohlfarth A, Auwärter V, Perdekamp MG, Wurst FM, Weinmann W. Urine tested positive for ethyl glucuronide and ethyl sulfate after the consumption of yeast and sugar. Forensic Sci Int 2010;202(1):e45-e7.
111. Politi L, Morini L, Groppi A, Poloni V, Pozzi F, Polettini A. Direct determination of the ethanol metabolites ethyl glucuronide and ethyl sulfate in urine by liquid chromatography/electrospray tandem mass spectrometry. Rapid Commun Mass Spectrom 2005;19(10):1321-31.
112. Brevik A, Veierød MB, Drevon CA, Andersen LF. Evaluation of the odd fatty acids 15:0 and 17:0 in serum and adipose tissue as markers of intake of milk and dairy fat. Eur J Clin Nutr 2005;59:1417.
113. Tasevska N, Midthune D, Potischman N, Subar AF, Cross AJ, Bingham SA, Schatzkin A, Kipnis V. Use of the Predictive Sugars Biomarker to Evaluate Self-Reported Total Sugars Intake in the Observing Protein and Energy Nutrition (OPEN) Study. Cancer Epidemiol, Biomarkers Prev 2011;20(3):490-500.
114. Tasevska N, Midthune D, Tinker LF, Potischman N, Lampe JW, Neuhouser ML, Beasley JM, Van Horn L, Prentice RL, Kipnis V. Use of a Urinary Sugars Biomarker to Assess Measurement Error in Self-Reported Sugars Intake in the Nutrition and Physical Activity Assessment Study (NPAAS). Cancer Epidemiol, Biomarkers Prev 2014;23(12):2874-83.
115. La Forge FB. D-Mannoketoheptose, a new sugar from the Avocado. J Biol Chem 1917;28(2):511-22.
116. Liu XS, James; Lu Arpaia, Mary; Madore, Monica A. Postulated Physiological Roles of the Seven-carbon Sugars, Mannoheptulose, and Perseitol in Avocado. J Am Soc Hortic Sci 2002;127(1):108-14.
117. Bean RC, Barr BK, Welch HV, Porter GG. Carbohydrate metabolism of the avocado. Arch Biochem Biophys 1962;96(3):524-9.

7 SUPPLEMENT

Supplemental Table S1.1. Overview of several definitions and classifications of biomarkers/markers used in nutritional research [50-53].

Classification	Definition
Based on function	*Biomarkers of dietary exposure*: assess dietary intake of different foods, nutrients, non-nutritive components or dietary patterns
	Biomarkers of nutritional status: reflect not only dietary intake but also metabolism of nutrients and possible effects from disease processes
	Biomarkers of health/disease: related to different intermediate phenotypes of a disease or even to the severity of the disease
Based on application	*Exposure/intake biomarker*: reflects level of extrinsic variables that humans are exposed to, such as diet or food compounds
	Food compound intake biomarkers: divided in nutrient and non-nutrient intake biomarkers, have distinctive dose- and time-dependent response after intake
	Food or food component intake biomarkers: measure intake of specific food groups, foods or food components (such as ingredients), have distinctive dose- and time-dependent response after intake
	Dietary pattern biomarkers: reflect the average diet of an individual, distinguish different dietary habits
	Effect biomarker: divided in functional response biomarkers and risk-effect biomarkers; monitor changes in biochemical, physiological or psychological state as a response to nutritional exposures
	Susceptibility biomarker: measurable indicator including genetic or acquired host factors and intrinsic factors, individual susceptibility or resilience to an exposure predicting intensity of its effect on the individual
	Food compound status biomarkers: divided in nutrient and non-nutrient intake biomarkers, reflect the current body burden or status of compounds from food
	Physiological or health state biomarkers: divided into host factor biomarkers (personal intrinsic quality or trait influencing susceptibility to disease) and risk (personal intrinsic quality or trait influencing susceptibility to develop a disease) biomarkers
Based on temporality	*Short-term/acute biomarkers*: reflecting intake over past hours/days
	Medium-term biomarkers: reflecting intake over weeks/months
	Long-term biomarkers: reflecting intake over months/years

Supplemental Table S1.1 continued.

Based on correlation with intake/ biokinetics	*Recovery biomarkers*: excretion is highly correlated with intake (not subject to homeostasis, no substantial inter-individual differences in metabolism) *Concentration biomarkers*: excretion is correlated with intake, affected by metabolism or personal characteristics (e.g. sex, age, smoking, obesity,..) *Replacement biomarkers*: serve as proxy for intake when it is not possible to capture intake due to limited nutrient databases to assess intake *Predictive biomarkers*: show dose-response relationship with intake levels, are sensitive, time-dependent, may be affected by personal characteristics (see above) but distinction is that overall recovery is lower

Supplemental Table S4.1. List of all used standard substances, including their purity and manufacturer. In a separate column additionally combination of standard substances and their concentrations for the used standard solutions are included.

FAME standards for RI			
Substance	**Purity**	**Manufacturer**	**FAME solution**
Methyl octanoate (C8)	GC	Merck, Darmstadt, Germany	Each FAME 250 µMol/L in heptane.
Methyl decanoate (C10)	99%	Sigma-Aldrich, Steinheim, Germany	
Methyl dodecanoate (C12)	99%	Sigma-Aldrich, Steinheim, Germany	
Methyl tetradecanoate (C14)	> 99%	Sigma-Aldrich, Steinheim, Germany	
Methyl hexadecanoate (C16)	> 99%	Sigma-Aldrich, Steinheim, Germany	
Methyl octadecanoate (C18)	> 99%	Merck, Darmstadt, Germany	
Methyl eicosanoate (C20)	> 99%	Sigma-Aldrich, Steinheim, Germany	
Methyl docosanoate (C22)	> 99%	Sigma-Aldrich, Steinheim, Germany	
Methyl tetracosanoate (C24)	> 99%	Sigma-Aldrich, Steinheim, Germany	
Methyl hexacosanoate (C26)	95%	Sigma-Aldrich, Steinheim, Germany	
Internal standards			
Substance	**Purity**	**Manufacturer**	**IS solution**
1-*O*-Methyl-2-desoxy-D-ribose	96%	Sigma-Aldrich, Steinheim, Germany	• 100 µMol/L
Phenyl-*β*-D-glucopyranoside	99%	ABCR, Karlsruhe, Germany	• 50 µMol/L
D-Pinitol	95%	Sigma-Aldrich, Steinheim, Germany	• 50 µMol/L in MeOH/dH_2O.
Sugar compounds			
Substance	**Purity**	**Manufacturer**	**Standard solutions (each 25 µMol/L)**
D-Threitol	99%	Sigma-Aldrich, Steinheim, Germany	Solution AA
Galactaric acid	97%	Sigma-Aldrich, Steinheim, Germany	
L-Threonic acid	>97%	Sigma-Aldrich, Steinheim, Germany	
Lacto-*N*-biose	-	Carbosynth, Berkshire, UK	
meso-Erythritol	-	Sigma-Aldrich, Steinheim, Germany	Solution AB
Glucono-*δ*-lactone	>99%	Sigma-Aldrich, Steinheim, Germany	
N-Acetyl-D-glucosamine	>99%	Carl Roth, Karlsruhe, Germany	
Lactose	99.5%	Sigma-Aldrich, Steinheim, Germany	
Ribitol	>99%	Sigma-Aldrich, Steinheim, Germany	Solution AC
Eryhtronolactone	>98%	TCI Chemicals, Tokyo, Japan	
Gluconic acid	>99%	Sigma-Aldrich, Steinheim, Germany	
N-Acetyl-D-mannosamine	>98%	Sigma-Aldrich, Steinheim, Germany	
Xylitol	>99%	Sigma-Aldrich, Steinheim, Germany	Solution AD
2,3-Dihydroxybutyric acid	-	Sigma-Aldrich, Steinheim, Germany	
Isosaccharinic acid-1,4-lactone	-	Carbosynth, Berkshire, UK	
N-Acetyl-D-galactosamine	>99%	Carl Roth, Karlsruhe, Germany	

Supplemental Table S4.1 continued.

D-Arabitol	>98%	TCI Chemicals, Tokyo, Japan	Solution AE
3,4-Dihydroxybutyric acid	-	Sigma-Aldrich, Steinheim, Germany	
α-D-Glucoheptonic acid	>99%	Sigma-Aldrich, Steinheim, Germany	
D-Ribose-5-phosphate	>99%	Sigma-Aldrich, Steinheim, Germany	
1,5-Anhydroglucitol	-	Sigma-Aldrich, Steinheim, Germany	Solution AF
2,4-Dihydroxybutyric acid	≥95%	Sigma-Aldrich, Steinheim, Germany	
D-Erythrose	≥75%	Sigma-Aldrich, Steinheim, Germany	
D-Fructose-6-phosphate	≥98%	Sigma-Aldrich, Steinheim, Germany	
L-Fucitol	-	Sigma-Aldrich, Steinheim, Germany	Solution AG
D-Erythrulose	≥85%	Sigma-Aldrich, Steinheim, Germany	
Glucose-1-phosphate	>98%	Carl Roth, Karlsruhe, Germany	
D-Mannitol	a.g.	Sigma-Aldrich, Steinheim, Germany	Solution AH
1-Deoxy-D-xylulose-5-phosphate	>99%	Sigma-Aldrich, Steinheim, Germany	
Glucose-6-phosphate	>98%	Merck, Darmstadt, Germany	
D-Sorbitol	-	Merck, Darmstadt, Germany	Solution AI
2-Desoxy-D-ribose	99%	Sigma-Aldrich, Steinheim, Germany	
Methyl-*β*-D-glucuronide	-	Carbosynth, Berkshire, UK	
Galactitol	>99%	Sigma-Aldrich, Steinheim, Germany	Solution AJ
D-Ribose	>98%	Carl Roth, Karlsruhe, Germany	
Methyl-*α*-D-glucopyranoside	99%	Sigma-Aldrich, Steinheim, Germany	
L-Iditol	-	Carbosynth, Berkshire, UK	Solution AK
D-Arabinose	>98%	Sigma-Aldrich, Steinheim, Germany	
Ethyl-*β*-D-glucuronide	-	Carbosynth, Berkshire, UK	
Perseitol	-	Carbosynth, Berkshire, UK	Solution AL
D-Xylose	r.g.	Sigma-Aldrich, Steinheim, Germany	
Ethylglucoside	-	Chemos GmbH, Altdorf, Germany	
Maltitol	97%	ABCR, Karlsruhe, Germany	Solution AM
D-Lyxose	99%	Sigma-Aldrich, Steinheim, Germany	
Lactulose	-	Carl Roth, Karlsruhe, Germany	
Galactinol	>98%	Sigma-Aldrich, Steinheim, Germany	Solution AN
D-Apiose	-	Carbosynth, Berkshire, UK	
Laminaribiose	-	Carbosynth, Berkshire, UK	
allo-Inositol	97%	Sigma-Aldrich, Steinheim, Germany	Solution AO
D-Ribulose	97%	Sigma-Aldrich, Steinheim, Germany	
Maltose	r.g.	Sigma-Aldrich, Steinheim, Germany	

Supplemental Table S4.1 continued.

chiro-Inositol	98%	ABCR, Karlsruhe, Germany	
D-Xylulose	-	Carbosynth, Berkshire, UK	Solution AP
Maltulose	-	Carbosynth, Berkshire, UK	
L-Fucose	>99%	Sigma-Aldrich, Steinheim, Germany	Solution AQ
Melibiose	-	Merck, Darmstadt, Germany	
myo-Inositol	-	Sigma-Aldrich, Steinheim, Germany	
L-Rhamnose	>99%	Merck, Darmstadt, Germany	Solution AR
D-Leucrose	-	Carbosynth, Berkshire, UK	
scyllo-Inositol	98%	ABCR, Karlsruhe, Germany	
Levoglucosan	99%	ABCR, Karlsruhe, Germany	Solution AS
Neohesperidose	-	Carbosynth, Berkshire, UK	
1-D-4-*O*-Methyl-*myo*-inositol		Carbosynth, Berkshire, UK	
D-Galactose	r.g.	Serva, Heidelberg, Germany	Solution AT
Neotrehalose	≥98%	Sigma-Aldrich, Steinheim, Germany	
D-Glucose	a.g.	Sigma-Aldrich, Steinheim, Germany	Solution AU
Nigerose	≥94%	Sigma-Aldrich, Steinheim, Germany	
cis-Inositol	-	Carbosynth, Berkshire, UK	
D-Mannose	>98.5%	Carl Roth, Karlsruhe, Germany	Solution AV
Palatinose	≥99%	Sigma-Aldrich, Steinheim, Germany	
epi-Inositol	-	Carbosynth, Berkshire, UK	
D-Allose	-	Carbosynth, Berkshire, UK	Solution AW
Rutinose	-	Carbosynth, Berkshire, UK	
neo-Inositol	-	Carbosynth, Berkshire, UK	
D-Altrose	-	Carbosynth, Berkshire, UK	Solution AX
Xylobiose	≥90%	Sigma-Aldrich, Steinheim, Germany	
2-Deoxy-D-ribonic acid-1,4-lactone	-	Carbosynth, Berkshire, UK	Solution AY
D-Gulose	-	Carbosynth, Berkshire, UK	
Turanose	98%	ABCR, Karlsruhe, Germany	
5-Deoxy-L-arabonic acid-1,4-lactone	-	Carbosynth, Berkshire, UK	Solution AZ
D-Psicose	>95%	Sigma-Aldrich, Steinheim, Germany	
Galactosamine	>97%	Carl Roth, Karlsruhe, Germany	
D-Arabino-1,4-lactone	-	Carbosynth, Berkshire, UK	
D-Fructose	99%	Sigma-Aldrich, Steinheim, Germany	Solution BA
α-Sophorose	-	Carbosynth, Berkshire, UK	

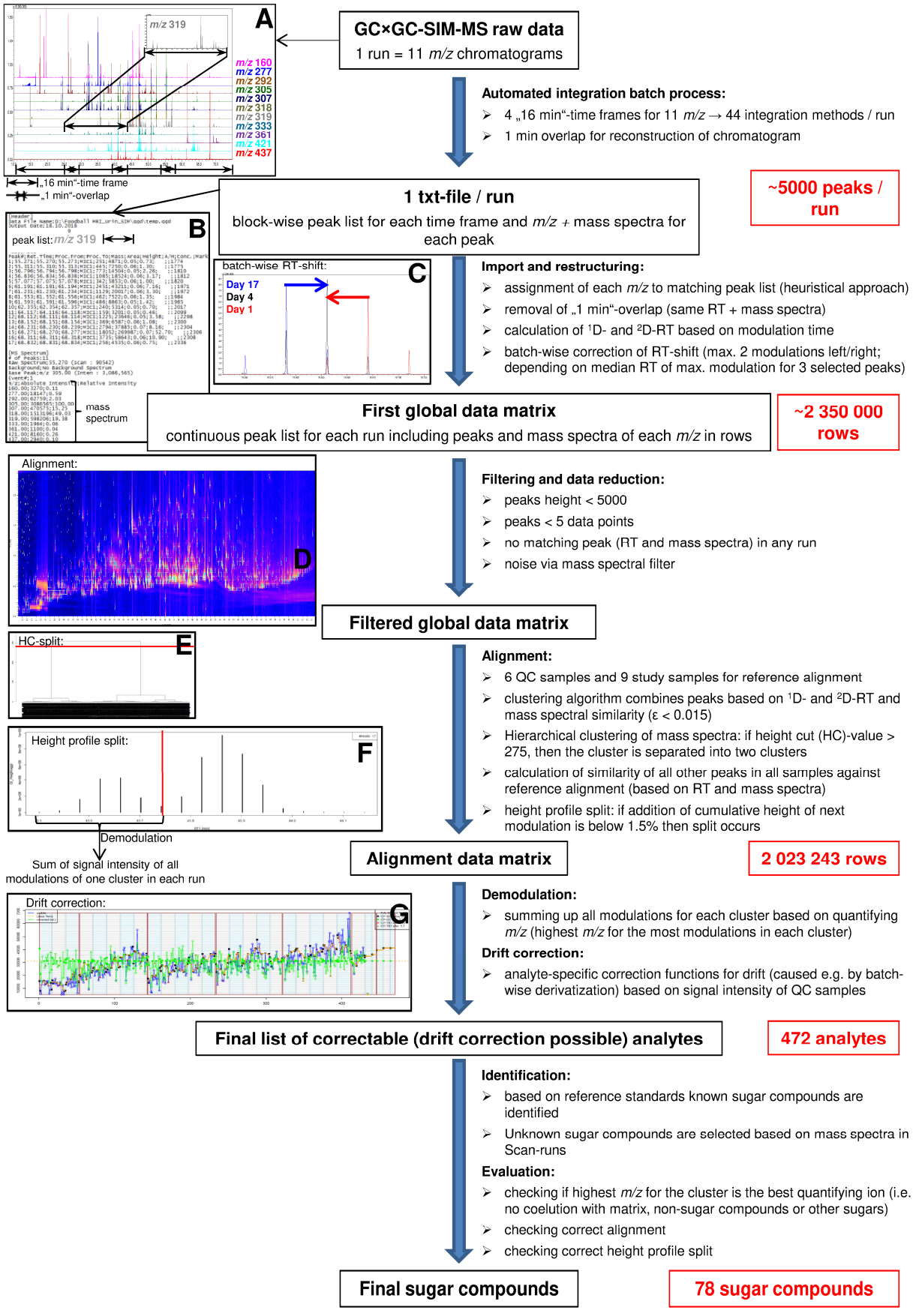
Supplemental Table S4.1 continued.

D-Ribono-1,4-lactone	95%	ABCR, Karlsruhe, Germany	
L-Sorbose	-	Sigma-Aldrich, Steinheim, Germany	Solution BB
β-Gentiobiose	96%	ABCR, Karlsruhe, Germany	
D-Xylonic acid	-	Carbosynth, Berkshire, UK	
D -Tagatose	99%	Sigma-Aldrich, Steinheim, Germany	Solution BC
D-Trehalose	>99%	Sigma-Aldrich, Steinheim, Germany	
D-Galacturonic acid	-	Merck, Darmstadt, Germany	
L-Idose	-	Carbosynth, Berkshire, UK	Solution BD
D-Cellobiose	>98%	Carl Roth, Karlsruhe, Germany	
D-Glucaric acid	>98.5%	Sigma-Aldrich, Steinheim, Germany	
D-Glycero-D-galacto-heptose	-	Carbosynth, Berkshire, UK	Solution BE
D-Sucrose	a.g.	Sigma-Aldrich, Steinheim, Germany	
D-Glucaric acid-1,4-lactone	-	Sigma-Aldrich, Steinheim, Germany	
D-Sedoheptulose	≥95%	Sigma-Aldrich, Steinheim, Germany	Solution BF
Epilactose	-	Carbosynth, Berkshire, UK	
D-Glucuronic acid	>98%	Carl Roth, Karlsruhe, Germany	
D-Mannoheptulose	-	Carbosynth, Berkshire, UK	Solution BG
Isomaltose	97%	ABCR, Karlsruhe, Germany	
D-Mannonic acid-1,4-lactone	-	Carbosynth, Berkshire, UK	
D-Glucosamine	>99%	Sigma-Aldrich, Steinheim, Germany	Solution BH
Isoprimeverose	-	Carbosynth, Berkshire, UK	
D-Galactono-1,4-lactone	-	Carbosynth, Berkshire, UK	
D-Mannosamine	>98%	Sigma-Aldrich, Steinheim, Germany	Solution BI
Kojibiose	-	Carbosynth, Berkshire, UK	

a.g.: analytical grade; r.g.: research grade

Supplemental Table S4.2. Consumables utilized in this study.

Consumable	Manufacturer
1.5 mL Reaction tubes (safe-lock™)	Eppendorf, Hamburg, Germany
Molecular sieve 0.3 nm, type 564	Carl Roth, Karlsruhe, Germany
Syringe, PTFE (25 µL, 50 µL)	Trajan Scientific, Ringwood, Australia
10 µL Syringe for CTC analytics systems, fixed needle, conical	Trajan Scientific, Ringwood, Australia
9 mm Screw thread vial 300 µL, fused insert, clear	Thermo Scientific, Langerwehe, Germany
9 mm Open top short screw AVCS cap, 6 mm hole, blue, Polypropylene, white silicone/Red PTFE	Thermo Scientific, Langerwehe, Germany
Pipette (0.1-10 µL, 0.5-20 µL, 10-100 µL, 2-200 µL, 50-1000 µL, 500-5000 µL)	Eppendorf, Hamburg, Germany
2.0 mL Crimp vial, amber	WICOM, Heppenheim, Germany
11 mm Aluminium crimp cap, Butylgummi/PTFE	WICOM, Heppenheim, Germany
16.0 mL Glass screw thread bottle with cap, Butylgummi/PTFE-septum	WICOM, Heppenheim, Germany



Supplemental Figure S4.1. Workflow of data processing.

Supplemental Table S4.3. Automated integration parameters.

Parameter	Setting/value
Integrated *m/z*	160.00, 277.00, 292.00, 305.00, 307.00, 318.00, 319.00, 333.00, 361.00, 421.00, and 437.00
Timeframe	Each ~ 16 min
Mode	Detail
Slope	2500 min^{-1}
Width	0.04 s
Drift	1000 min^{-1}
T.DBL	1000 min
Min Height	3000
Smooting	None
Spectrum process (Raw spectrum)	Peak Top Spectrum/apex
Spectrum process (Background spectrum)	None

Supplemental Table S4.4. Monosaccharides detected in human urine with the one-dimensional GC-MS method and the two-dimensional GC×GC-MS method.

Sugar compounds	GC-MS	GC×GC-MS
Apiose		X[1]
Xylose	X	X
Arabinose	X	X
Ribose	X	X
Ribulose/Xylulose		X[2]
Fucose	X	X
Rhamnose		X
Tagatose		X
Psicose	X	X
Fructose	X	X
Allose	X	X
Galactose	X	X
Glucose	X	X
Mannose	X	X
Sorbose		Traces[3]
Mannoheptulose	X	X
Unknown monosaccharide A0986_H2		X
Sedoheptulose	X	X
Unknown monosaccharide U03/A0408	X	X
Unknown monosaccharide U21/A1005_H1	X	X

[1] no drift correction; [2] not separable, could be either or both; [3] identification unsure due to low signal intensity

Supplemental Table S4.5. Disaccharides detected in human urine with the one-dimensional GC-MS method and the two-dimensional GC×GC-MS method.

Sugar compounds	GC-MS	GC×GC-MS
Sucrose	X	X
Lactose	X	X
Maltose	X	X
Isomaltose		Traces[3]
Unknown disaccharide U26/A1254	X	X
Unknown disaccharide U27/A1307	X	X
Unknown disaccharide U28	X	
Unknown disaccharide U29	Traces	
Unknown disaccharide A1235		X
Unknown disaccharide A1283		X
Unknown disaccharide A1289		X
Unknown disaccharide A1302		X

[3] identification unsure due to trace amounts

Supplemental Table S4.6. Polyols detected in human urine with the one-dimensional GC-MS method and the two-dimensional GC×GC-MS method.

Sugar compounds	GC-MS	GC×GC-MS
Threitol	X	X
meso-Erythritol	X	X
1-Deoxypentitol A0220		X
1-Deoxypentitol A0226		X
2-Deoxypentitol		X
Unknown polyol U02/A0402	X	X
Xylitol	X	X
Arabitol	X	X
Ribitol	X	X
1-Deoxyhexitol A0494_H1	X	X
1-Deoxyhexitol A0494_H2_H1		X
1,5-Anhydroglucitol		X
Fucitol	X	X
Mannitol	X	X
Sorbitol	X	X
Galactitol	X	X
Perseitol	X	X
chiro-Inositol	X	X
scyllo-Inositol	X	X
myo-Inositol	X	X
allo-Inositol		Traces[4]
cis-Inositol		Traces[4]

[4] trace amounts, no analyte cluster detected

Supplemental Table S4.7. Sugar acids detected in human urine with the one-dimensional GC-MS method and the two-dimensional GC×GC-MS method.

Sugar compounds	GC-MS	GC×GC-MS
Glyceric acid		X
4-Deoxyerythronic acid		X
4-Deoxythreonic acid		X
2-Deoxytetronic acid		X
Erythronic acid	X	X
Threonic acid	X	X
Deoxy sugar acid A0250		X
2-Deoxyribonic acid/Unknown sugar acid U01	X	X
Deoxy sugar acid A0279_H1		X
Deoxy sugar acid A0279_H2_H1		X
5-Deoxyarabonic acid		X
Unknown sugar acid U06/A0471	X	X
Xylonic acid	X	X[5]
Ribonic acid	X	X[5]
Unknown sugar acid U09/A0510	X	X
Arabonic acid	X	X
Unknown sugar acid A0532		X
Unknown sugar acid A0564		X
Glucuronic acid	X	X
Mannonic acid	X	Traces
Galactonic acid	X	X
Glucaric acid		X
Gluconic acid	X	X

[5] coeluting in the GC×GC-MS method.

Supplemental Table S4.8. Other sugar compounds and unknown sugar-like compounds detected in human urine with the one-dimensional GC-MS method and the two-dimensional GC×GC-MS method.

Sugar compounds	GC-MS	GC×GC-MS
Levoglucosan	X	X
Ethyl-*β*-D-glucuronide	X	Traces[4]
Methylglucopyranoside		Traces[4]
N-Acetyl-D-glucosamine		X[5]
N-Acetyl-D-mannosamine		X[5]
N-Acetyl-D-galactosamine		X
Glucosamine		X
Unknown U05/A0461_H1	X	X
Unknown U12/A0569	X	X
Unknown U24	X	
Unknown U25	X	
Unknown A0286		X
Unknown A0664		X

[4] trace amounts, no analyte cluster detected; [5] coeluting in the GC×GC-MS method.

Supplemental Table S4.9. All measured sugar compounds are listed together with their inter- and mean intra-day CV, median peak height, frequency in samples (Pr%) and QC's (QC%), quantifying ion and the occurrence of coelution or partial coelution. Additionally, sugar compounds that only occur in traces or were saturated are marked.

Sugar compounds	Corrected		Uncorrected		median peak height	Pr%	QC%	m/z	Coelution
	inter-day CV	mean intra-day CV	inter-day CV	mean intra-day CV					
Apiose[1]	12.6		12.6		1.17E+04	33.0	22.9	307	
Xylose deriv. 1	4.9	3.7	22.0	8.1	5.53E+05	100	100	307	
Xylose deriv. 2	4.5	3.8	17.8	6.8	2.94E+06	100	100	307	partial coelution
Arabinose	5.0	4.0	21.5	8.2	6.36E+06	100	100	307	
Ribose	5.3	4.3	21.5	8.0	2.48E+06	100	100	307	
Ribulose/Xylulose	6.4	5.4	22.7	11.3	8.10E+04	100	100	333	coelution
Fucose	5.3	4.4	21.3	8.4	5.78E+05	100	100	277	
Rhamnose	5.6	4.8	20.5	8.3	1.50E+05	100	100	160	
Tagatose[1]	6.0	5.3	25.4	10.8	3.46E+04	98.1	99.3	307	
Psicose deriv. 1	4.7	3.6	23.4	9.0	1.10E+06	100	100	307	
Psicose deriv. 2	5.1	4.2	23.4	9.2	1.88E+06	100	100	307	
Fructose deriv. 1	4.9	4.2	23.7	9.4	8.32E+05	100	100	307	
Sorbose	6.1	5.2	26.6	15.4	7.74E+04	100	100	307	
Fructose deriv. 2	5.3	4.3	23.1	9.4	5.49E+05	100	100	307	
Allose deriv. 1	4.8	4.0	22.8	8.3	2.73E+06	100	100	319	
Mannose deriv. 1	5	4.1	23.1	8.5	6.31E+05	100	100	319	
Galactose deriv. 1	5.1	4.1	22.1	9.0	7.99E+05	100	100	319	
Glucose deriv. 1[2]	1.9	1.6	8.4	3.6	1.53E+07	100	100	319	
Mannose deriv. 2/Allose deriv. 2	5.3	4.5	22.3	8.7	8.26E+04	100	100	319	coelution
Galactose deriv. 2/Glucose deriv. 2	5.8	5.1	22.3	8.4	1.80E+05	100	100	319	coelution
Mannoheptulose	5.0	4.1	26.6	11.6	7.56E+05	100	100	319	
Unknown monosaccharide A0986_H2	6.4	5.8	24.6	10.9	1.50E+06	100	100	319	

Supplemental Table S4.9 continued.

Sedoheptulose	5.4	4.7	26.6	9.5	7.86E+06	100	100	319	
Unknown monosaccharide U03/A0408	4.9	3.9	23.3	8.0	9.14E+05	100	100	307	
Unknown monosaccharide U21/A1005_H1	5.9	5.3	26.6	9.9	6.98E+05	100	100	319	
Sucrose	6.1	5.4	53.9	16.7	8.24E+05	100	100	361	
Lactose deriv. 1	6.4	5.3	26.6	11.7	7.55E+05	100	100	361	
Lactose deriv. 2	7.1	6.2	24.9	11.4	1.70E+05	100	100	361	partial coelution
Maltose deriv. 1	6.6	5.5	28.9	13.0	3.47E+05	100	100	361	coelution
Maltose deriv. 2	5.5	4.8	26.2	10.9	2.09E+05	100	100	361	coelution
Isomaltose[1]	9.4	8.1	36.6	17.0	3.45E+04	95.5	100	361	
Unknown disaccharide U26/A1254	9.1	7.0	28.0	13.0	2.18E+05	79.2	99.3	319	
Unknown disaccharide U27/A1307	6.5	5.6	20.2	9.9	1.93E+05	100	100	305	
Unknown disaccharide A1235[1]	20.2	19.8	44.0	25.3	3.41E+04	15.9	83.3	319	partial coelution
Unknown disaccharide A1283	7.2	6.5	29.7	13.0	3.78E+05	99.6	100	319	partial coelution
Unknown disaccharide A1289	8.1	7.8	35.4	15.7	8.37E+04	73.1	99.3	319	
Unknown disaccharide A1302	9.4	7.9	31.1	14.7	8.49E+04	99.2	100	361	
Threitol	4.9	3.9	24.5	10.1	1.41E+06	100	100	307	
meso-Erythritol	4.7	3.9	22.4	6.3	5.98E+06	100	100	307	
1-Deoxypentitol A0220	5.4	4.6	21.6	8.0	5.15E+05	100	100	307	
1-Deoxypentitol A0226	4.8	3.9	21.5	8.0	6.97E+05	100	100	307	
2-Deoxypentitol[1]	6.7	5.8	25.9	11.6	3.96E+04	100	100	307	
Unknown polyol U02/A0402	5.2	4.5	24.0	8.7	1.77E+05	100	100	307	
Xylitol	5.5	4.3	22.9	7.9	1.48E+06	100	100	307	
Arabitol	4.9	4.4	22.0	7.0	8.11E+06	100	100	307	
Ribitol	5.2	4.7	22.3	8.1	2.27E+06	100	100	319	
1-Deoxyhexitol A0494_H1	7.7	6.2	45.8	13.8	3.37E+05	100	97.9	319	partial coelution
1-Deoxyhexitol A0494_H2_H1	5.6	4.6	24.2	8.5	9.59E+05	100	100	319	partial coelution
1,5-Anhydroglucitol	6.6	6.1	22.9	10.2	6.09E+04	100	100	319	

Supplemental Table S4.9 continued.

Fucitol	6.0	5.4	23.3	8.8	3.09E+05	100	100	319	
Mannitol[2]	2.5	2.0	9.9	4.7	2.13E+07	100	100	319	
Sorbitol	5.8	5.0	22.4	8.2	5.74E+06	100	100	319	
Galactitol	5.6	4.9	23.2	8.1	1.47E+06	99.6	100	307	
Perseitol[1]	30.6		53.0		1.11E+04	48.1	86.1	319	
chiro-Inositol	4.5	3.3	22.9	7.5	3.58E+05	100	100	318	
scyllo-Inositol[2]	5.4	4.6	21.5	7.8	1.07E+07	100	100	318	
myo-Inositol[2]	3.9	3.7	15.5	7.0	1.29E+07	100	100	305	
epi-Inositol	5.4	4.7	21.7	8.5	1.56E+05	100	100	305	
neo-Inositol	10.6	6.7	23.4	10.7	1.98E+06	100	100	318	
Glyceric acid	5.0	4.5	16.1	7.9	5.87E+05	100	100	292	
4-Deoxyerythronic acid	3.4	2.9	15.4	6.5	6.43E+06	100	100	292	
4-Deoxythreonic acid[2]	2.2	1.8	10.4	6.0	1.33E+07	100	100	292	
2-Deoxytetronic acid	6.6	6.1	24.0	11.8	5.63E+04	98.5	100	305	partial coelution
Erythronic acid[2]	1.6	1.4	6.9	3.7	2.01E+07	100	100	292	
Threonic acid	4.9	4.1	18.6	7.1	8.67E+06	100	100	292	
Deoxy sugar acid A0250	5.3	4.5	19.2	8.0	8.64E+05	100	100	292	
2-Deoxyribonic acid/Unknown sugar acid U01	4.9	4.1	20.8	7.9	7.20E+05	100	100	333	partial coelution
Deoxy sugar acid A0279_H1	10	8.3	36.6	20.9	1.85E+04	93.5	100	292	
Deoxy sugar acid A0279_H2_H1	7.4	6.5	25.1	15.4	2.45E+04	99.6	100	292	
5-Deoxyarabonic acid	4.5	3.5	20	7.9	1.02E+05	100	100	292	
Sugar acid A0471/Unknown sugar acid U06	6.3	5.9	21.7	9.2	1.90E+05	100	100	292	
Xylonic acid/Ribonic acid	5.5	4.6	21.1	8.5	4.62E+06	100	100	292	coelution
Sugar acid A0510/Unknown sugar acid U09	5.2	4.2	21.3	7.7	1.53E+06	100	100	292	
Arabonic acid	5.0	4.2	22.1	8.2	5.25E+06	100	100	292	
Sugar acid A0532[1]	40.7	41.0	63.6	50.5	1.22E+04	61.4	81.3	292	partial coelution

Supplemental Table S4.9 continued.

Sugar acid A0564	5.6	4.6	22.9	9.2	1.28E+05	100	100	292	
Glucuronic acid deriv. 1[2]	3.6	3.1	12.9	5.1	1.85E+07	100	100	333	
Glucuronic acid deriv. 2	5.2	4.1	21.2	8.2	2.92E+06	100	100	333	
Mannonic acid[1]	7.5	6.7	26.5	12.6	3.39E+04	98.9	100	305	
Galactonic acid[1]	23.7	23.2	40.1	29.1	9.32E+03	41.7	86.8	305	partial coelution
Glucaric acid	6.3	5.1	24.9	10.0	2.44E+06	100	100	333	
Gluconic acid	5.2	4.2	22.6	7.9	9.50E+06	100	100	333	
Glucono-*δ*-lactone	6.4	5.5	21.2	9.4	1.98E+06	100	100	319	
Glucono-*γ*-lactone	5.3	5.0	24.2	10.8	7.38E+04	100	100	305	
Levoglucosan	4.9	4.1	22.3	9.7	8.69E+05	100	100	333	
N-Acetyl-D-glucosamine/*N*-Acetyl-D-mannosamine	5.7	4.5	25.6	9.4	6.80E+05	100	100	319	coelution
N-Acetyl-D-galactosamine	5.8	4.7	25.1	10.1	4.08E+05	100	100	319	
Glucosamine	26.4	22.6	48.8	32.7	7.42E+04			160	partial coelution
Unknown U05/A0461_H1	4.1	3.5	20.1	6.2	3.46E+06	100	100	305	
Unknown U12/A0569	5.3	4.2	24.0	8.3	5.51E+05	100	100	307	
Unknown A0286[1]	23.6	16.1	47.0	27.2	2.14E+04	34.1	92.4	307	
Unknown A0664[1]	8.8	8.1	29.5	15.9	2.19E+04	66.7	99.3	319	partial coelution
Methyl-desoxyribose deriv. 1	3.3	2.4	13.8	5.1				160	
Methyl-desoxyribose deriv. 2	2.2	1.7	11.0	4.5				160	
Pinitol	3.0	2.0	19.5	4.3				318	
Phenylglucopyranoside	10.5	6.9	32.0	12.0				319	

[1] Trace amount (peak height < 5.00E+04); [2] saturated analyte (peak height > 1.00E+07)

Supplemental Table S4.10. Data points of 15 selected sugar compounds for both the one-dimensional GC-MS sugar profiling method and the two-dimensional GC×GC-MS method.

Sugar compounds	GC-MS	GC×GC-MS
Arabinose	15	7
Fucose deriv. 2	11	6
Fructose deriv. 2	11	7
Glucose deriv. 1	12	7
Sedoheptulose deriv. 1	14	5
Sucrose	13	9
Lactose deriv. 1	11	8
Threitol	10	6
Ribitol	10	5
Sorbitol	10	6
scyllo-Inositol	12	7
Erythronic acid	9	8

Supplemental Table S5.1. Application of the validation scheme by Dragsted *et al.* [74] on the putative avocado consumption markers mannoheptulose and perseitol (section 3.2).

Validation questions [74]	**Avocado – mannoheptulose and perseitol**	
1. Plausibility: Is the marker compound plausible as a specific biomarker of food intake for the food or food group?	Yes	- both occur naturally in avocados (sugar profiling of 75 fruit and vegetables section 3.2 and literature [115-117])
- variability of (parent) compound in food or food group is limited		- only one variety (Hass avocado) typically consumed, but concentrations may vary depending on ripening stage [116]
- level of compound in other food is comparatively low or presence only in other foods not commonly consumed		- mannoheptulose occurs in other fruit and vegetables, perseitol only in traces in a few fruit and vegetables (section 3.2)
2. Dose-response: Is there a dose-response relationship at relevant intake levels of the targeted food (quantitative aspect)?	Yes	- dose-response relationship not yet evaluated with respect to quantitative consumption of avocado
- established using several intake levels (meal study) or different meal studies with comparable results or cross-sectional or longitudinal observational study		- qualitative dose-response established in cross-sectional observational study (section 3.2) and in a meal study with several intake levels [109]
- background level is 0 or low		- low for mannoheptulose; traces for perseitol (section 3.2)
3. Time-response: Is the biomarker kinetics described adequately to make a wise choice of sample type, frequency and time window?	Yes	- kinetics in urine see [109], excretion within 16 hours after consumption: short-term or acute marker
- single-meal time response relationship has been described for defined sample type and time window in a meal study		- yes, see above
- kinetics after repeated intakes has been described for defined sample type in a meal study or accumulation in certain sample types has been observed		- no, however, both sugars are highly polar and as such unlikely to accumulate in the human body

Supplemental Table S5.1 continued.

4. Robustness: Has the marker been shown to be robust after intake of complex meals reflecting dietary habits of the targeted population?	Yes	- KarMeN study is an observational cross-sectional study with participants on an unrestricted and diverse diet (section 3.2)
- robust after intake of complex meals in intervention studies or observational studies		- intervention study not available; observational study (section 3.2)
- no confounding food observed or its level is low or confounding foods are not commonly consumed		- no confounding food observed for perseitol (section 3.2), minor confounding for mannoheptulose possible (occurs in other food as well; section 3.2)
5. Reliability: Has the marker been shown to compare well with other markers or questionnaire data for the same food/food group?	partially[1]	- in KarMeN study self-reported intake of a 24 hour recall was used to establish correlation (section 3.2), but no objective comparison between the 24 hour recall and the marker was performed
- marker has compared well with other biomarkers of same food or food group or with dietary assessment instruments or with data in studies with highly controlled setting and supervised intake		- no other markers known and no studies with controlled setting available
6. Stability: Is the marker chemically and biologically stable during biospecimen collection and storage, making measurements reliable and feasible?	partially	- no data available
7. Analytical performance: Are analytical variability (CV%), accuracy, sensitivity and specificity known as adequate for at least one reported analytical method?	Yes	- see section 3.2 for KarMeN study and [109]
- protocol of method has been well described and could be repeated		- see section 3.2 for KarMeN study and [109]
- method has been compared with validated method or reference		- no data available
8. Reproducibility: Has the analysis been successfully reproduced in another laboratory?	partially	- no data available
		- however, at least two laboratories with different methods identified both sugars as potential dietary marker (see section 3.2 and [109]

[1] partially: not enough or no data available to fully answer with yes or no.

8 DANKSAGUNG

Mein herzlichster Dank gilt zuerst meiner Doktormutter, Prof. Sabine E. Kulling, für die Bereitstellung der äußerst interessanten, interdisziplinären und herausfordernden Aufgabenstellung, das in mich gesetzte Vertrauen und die fortwährende Unterstützung mit vielen konstruktiven Denkanstößen.

Ein besonderes Danke an dieser Stelle auch an Prof. Wolfgang Schwack für die freundliche Mitbetreuung meiner Arbeit von Seiten der Universität Hohenheim und an Prof. Walter Vetter für die spontane Übernahme der Mitbetreuung meiner Dissertation und die Erstellung des Zweitgutachtens. Außerdem möchte ich auch Prof. Lutz Graeve für die Bereitschaft danken, als 3. Prüfer für meine Dissertation zur Verfügung zu stehen.

Mein ganz besonderer Dank gilt Dr. Christoph Weinert für die gute Zusammenarbeit, die uneingeschränkte Unterstützung auf allen Gebieten, für die vielen konstruktiven Anregungen und Ideen, sowie für seine große Hilfsbereitschaft bei jeglichen Problemen technischer, statistischer oder auch sonstiger Natur.

An dieser Stelle danke ich auch Prof. Chiara Cordero, Prof. Carlo Bicchi und Dr. Erica Liberto für die Möglichkeit des dreimonatigen Aufenthalts an der Universität Turin und die Chance, dort viel über ein anderes GC×GC-qMS System im Hinblick auf volatile Verbindungen und die bildbasierte Prozessierung von GC×GC-qMS Daten zu lernen.

Ganz besonders bedanken möchte ich mich bei Prof. Hannelore Daniel und den Mitarbeitern des Studienzentrums der Technischen Universität München für die Bereitstellung der Plasma Proben der MIPROMET Studie und die sehr hilfreichen Denkanstöße zu vielen physiologischen Zusammenhängen.

Weiterhin danke ich Herrn Björn Egert für die gewinnbringenden, kreativen Diskussionen über die Anpassung der peakbasierten Prozessierung auf GC×GC-SIM-MS Daten und für seine unermüdliche Programmierarbeit für die Umsetzung dieser Ideen.

Ebenso möchte ich Prof. Bernhard Watzl, Prof. Achim Bub und den Mitgliedern des Studienzentrums des MRI für die Durchführung und Bereitstellung der KarMeN-Proben danken. Prof. Ingrid Hoffmann, Dr. Eva Hummel, Dr. Friederike Wittig und

dem Institut EV möchte ich für die Erhebung der Verzehrsdaten im Rahmen der KarMeN-Studie danken.

Bei Silvia Remmert, Claudia Hoffmann und Michael Meyer aus dem GC-Labor und den Mitgliedern der Sensorik-Runde möchte ich mich herzlich für die sehr gute und zuverlässige Unterstützung bei allerlei praktischen Arbeiten, für ihre Kollegialität, den humorvollen Umgang miteinander, die Unternehmungen, die netten Pausengespräche und das Mitfiebern bedanken. Auch danke ich allen Mitgliedern des Meta-Kolloquiums, die zum Gelingen dieser Arbeit beigetragen haben.

Zu großem Dank bin ich meinen Mitstreitern und meinen guten Freunden Nicolas Danylec, Dr. Falco Beer, Freya Ziegler, Mona Ehlers und Franziska Scharmann verpflichtet, für ihre unermüdliche Unterstützung und Kollegialität, für ihre Bereitschaft für Diskussionen, die vielen sonstigen gemeinsamen Unternehmungen und ihren Humor.

Mein besonderer Dank geht an meine Eltern und meinen Bruder, dafür, dass sie mich in den letzten Jahren immer in allen Lebenslagen unterstützt und ermutigt haben und mir dadurch diesen Weg erst ermöglicht haben.

“Expectation wasn’t just about what people expected of you.

It was about what you expected of yourself.”

– Shallan Davar in Words of Radiance by Brandon Sanderson

9 LEBENSLAUF

PERSÖNLICHE ANGABEN

Name:	Carina I. Mack
Geburtsdatum/-ort:	27.11.1989 in Ostfildern Ruit
Staatsangehörigkeit:	deutsch

WISSENSCHAFTLICHE TÄTIGKEIT

seit 02/2015	Wiss. Mitarbeiterin am Max Rubner-Institut Karlsruhe, Institut für Sicherheit und Qualität bei Obst und Gemüse (Prof. Dr. Sabine Kulling)
09/2016 - 12/2016	Forschungsaufenthalt an der Universität Turin (Prof. Dr. Chiara Cordero)

HOCHSCHULAUSBILDUNG

07/2014 - 01/2015	Diplomarbeit am Max Rubner-Institut Karlsruhe, Institut für Sicherheit und Qualität für Obst und Gemüse zum Thema: „Der Reifungsverlauf von Kiwis (*Actinidia deliciosa* cv. Hayward): Ungerichtete Metabolomanalyse mittels GC×GC-MS und Entwicklung einer GC-MS-Profilingmethode für Zuckerverbindungen“
10/2011 - 01/2015	Hauptstudium der Lebensmittelchemie an der Universität Hohenheim
10/2009 - 09/2011	Grundstudium der Lebensmittelchemie an der Universität Stuttgart

SCHULAUSBILDUNG

09/2006 - 07/2009	Allgemeine Hochschulreife an der Käthe-Kollwitz-Schule in Esslingen
09/2000 - 07-2006	Mittlere Reife an der Immanuel-Kant-Realschule in Leinfelden-Echterdingen
09/1996 - 07/2000	Grundschule in Leinfelden-Echterdingen