Whole grains are an abundant source of phytochemicals, which contribute to the health benefits of plant-based foods. This study showed that enzymatic bioprocessing, sourdough fermentation, and colonic microbial metabolism have a major impact on the abundance of whole grain phytochemicals, affecting their bioavailability and even producing novel metabolites, including betainized compounds. The study further supports the applicability of metabolomics in food science and nutrition.
EFFECTS OF FOOD PROCESSING AND GUT MICROBIAL METABOLISM ON WHOLE GRAIN PHYTOCHEMICALS

A METABOLOMICS APPROACH
Ville Mikael Koistinen

EFFECTS OF FOOD PROCESSING AND GUT MICROBIAL METABOLISM ON WHOLE GRAIN PHYTOCHEMICALS

A METABOLOMICS APPROACH
“Life is not easy for any of us. But what of that? We must have perseverance and above all confidence in ourselves. We must believe that we are gifted for something and that this thing must be attained.”

Marie Skłodowska-Curie (1867–1934)

For my mother Raija (1959–2008)
ABSTRACT

Consumption of whole grains is associated with a decreased risk of all-cause mortality and several major chronic diseases. This protective effect of whole grains has been attributed to the high content of dietary fibre in the bran section and the wide array of phytochemicals, plant secondary metabolites with potential health effects, which exist within the fibre matrix. The mechanism of action of phytochemicals in disease prevention is largely unknown, but it is likely additive and synergistic, i.e. the compounds need to be ingested together to exert their antioxidative and other physiological effects. Before dietary phytochemicals reach their site of action in the body, they are released from the fibre matrix mainly during food processing and by gut microbiota. Simultaneously they are exposed to metabolism, altering the structure of the compounds entering the circulation.

The aim of this doctoral thesis was to elucidate the role of enzymatic bioprocessing, sourdough fermentation, and in vitro gut microbial metabolism on the overall metabolite profile of whole-grain wheat and rye. In specific, the thesis studies the transformations of two major phytochemical classes of whole grains, betaines and phenolic acids. Nontargeted metabolic profiling utilizing liquid chromatography coupled with quadrupole time-of-flight (qTOF) mass spectrometry was the main method applied to detect and identify the phytochemicals and their metabolites.

The results presented in this thesis confirm and expand previous findings on the effect of bioprocessing and gut microbial metabolism on phenolic acids. They also include novel findings, such as the unique metabolic profile of sourdough-fermented whole-grain rye and wheat bread, including phenolic acid and amino acid metabolites and potentially bioactive small peptides. Several metabolites, such as certain amino acid-derived betaines, were characterized in cereals for the first time. Evidence was obtained on how cereal bran-enriched diets alter the gut microbial composition in mice and how microbiota contributes to the metabolism of glycine betaine into other betainized compounds.
In conclusion, this thesis demonstrates the applicability of nontargeted metabolomics in 1) studying complex changes in the metabolite profiles of whole-grain cereals induced by food processing and microbial metabolism and 2) creating new hypotheses for targeted mechanistic studies on transformations and physiological effects of phytochemicals. The thesis highlights the potential of food processing, such as sourdough fermentation, in creating functional foods with increased abundance of health-promoting compounds, and the importance of gut microbiota in mediating the health effects.

*Keywords: bioprocessing, gut microbiota, mass spectrometry, metabolomics, phytochemicals, sourdough, whole grains*
Tiivistelmä

Täysjyväviljien kulutus on yhdistetty kuolleisuuden ja merkittävien kroonisten sai-
rauksien vähentyneeseen riskiin. Suojaavan vaikutuksen taustalla on arveltu olevan
täysjyvän leseen korkea ravintokuitupitoisuus sekä laaja joukko fytokemikaaleja,
kasvien sekundäärimetabolitiitteja, joilla on useita mahdollisia terveysvaikutuksia.
Fytokemikaalien vaikutusmekanismi sairauksien ehkäisyssä on suurelta osin tunte-
maton, mutta se on luonteleltaan additiivinen ja synergiitä, eli yhdisteelä on saa-
tava elimistöön yhtäaikaisesti havaitun vaikutuksen aikaansamiseksi. Ennen kuin
fytokemikaalit saavuttavat kohteensa elimistössä, niiden on vapauduttava kuitumat-
riisista pääasiassa ruokaprosessoinnin ja suoliston mikrobifermentaation seurauk-
sena. Samalla ne altistuvat metabolialle, joka muuttaa yhdisteiden rakennetta ennen
kuin ne imeytyvät verenkiihtoon.

Tämän väitöskirjatyön tavoitteena oli selvittää entsymaattisen bioprosessoinnin,
raskituksen (taikinajuurimenetelmän) ja *in vitro*-suolistomikrobifermentaation vai-
kutus täysjyvävehnän ja -rukoihin metabolityyppiliinin. Erityisesti työssä päädyttiin
uttamaan ja ymmärtää täysjyväviljien kannalta tärkeää fytokemikaaliryhmää, betaaiineja
ja fenoliapioita. Keskeisenä tutkimusmetodina työssä käytettiin nestekromatogra-
fi–qTOF-massaspektrometriaa hyödyntävää kohdentamatonta metabolityyppilii-
ointia.

Jäljempänä esiteltävät tulokset vahvistavat ja laajentavat aiempia löytyöksiä bio-
prosessoinnin ja suolistovermentoinnin vaikutuksista fenoliapioihin. Ne sisältävät
myös uusia havaintoja, kuten raskitetun täysjyvävehnän- ja ruisleivän ainutlaatuisen
metabolityyppiliinin, jolle tyyppiliinä ovat fenoli- ja aminohappojen metabolitiit sekä
pienten peptidin, joilla on potenssiäalisia terveysvaikutuksia. Useita metabolitiittejä, kuten
aminohaposta johdettuja betaaiineja, havaittiin ensimmäistä kertaa viljoista. Li-
säksi saatiin näyttää siitä, millä tavoin leseetä sisältävyrukovalijo muuttaa suoliston
mikrobioton koostumusta hiirillä ja että mikrobit osallistuvat glysinibetaainin meta-
boliaan muoksi betaainihdisteiksi.
Yhteenvetona tämä väitöskirjatyö vahvistaa kohdentamattoman metabolomiikan soveltumisen 1) ruokaprosessoinnin ja mikrobimetabolian aiheuttamien monimuutkaisten metabolisten muutosten tutkimiseen täysjyväviljoissa ja 2) luomaan uusia tutkimushypoteeseja kohdennetuille mekanistisille tutkimuksille fytokemikaalien rakennemuutoksista ja fysiologisista vaikutuksista. Väitöskirja korostaa ruokaprosessoinnin potentiaalia sellaisten funktionaalisten elintarvikkeiden kehittämisessä, jotka sisältävät enemmän terveyttä edistäviä yhdisteitä, ja suolistomikrobiston merkitystä terveysvaikutusten välittäjänä elimistöön.

Avainsanat (Yleinen suomalainen asiasanasto): aineenvaihduntatuotteet; bioaktiiviset yhdisteet; hapattaminen; fermentointi; massaspektrometria; nestekromatografia; ruis; suolistomikrobisto; täysjyväjauhot
Konsumtion av hela korn är associerad med en minskad risk för dödlighet och flera allvarliga kroniska sjukdomar. Denna skyddande effekt av hela korn har tillskrivits det höga innehållet av kostfiber i klisektionen och det stora utbudet av fytokemikalier, sekundära växtmetaboliter med potentiella hälsoeffekter, som finns inom fibermatrisen. Verkningsmekanismen för fytokemikalier vid sjukdomsprevention är för det mesta okänd, men det är sannolikt additiv och synergistisk, d.v.s. föreningarna måste intagas tillsammans för att utöva sina antioxidativa och andra fysiologiska effekter. Innan fytokemikalier i kosten når sin verkningsplats i kroppen, släpps de från fibermatrisen huvudsakligen under matförädling och genom tarmmikrobiella aktiviteter. Samtidigt utsätts de för metabolismen och förändrar strukturen hos de föreningar som kommer in i cirkulationen.

Syftet med denna doktorsavhandling var att belysa rollen som enzymatisk bioprocessering, surdjursjäsning och in vitro tarmmikrobiell metabolism på den övergripande metabolitprofilen av fullkornsvete och råg. Specifikt studerar avhandlingen omvandlingarna av två viktiga klasser av fytokemikalier i hela korn, betainer och fenolsyra. Örtkad metabolisk profilering med användning av vätskekromatografi kopplad till qTOF-masspektrometri var den huvudsakliga metoden som användes för att detektera och identifiera fytokemikalierna och deras metaboliter.

Sammanfattningsvis visar denna avhandling tillämpligheten av oriktad metabolomik i 1) studier av komplexa förändringar i metabolitprofilerna av helkornsprodukter som induceras av livsmedelsbearbetning och mikrobiell metabolism och 2) att skapa nya hypoteser för riktade mekanistiska studier om transformationer och fysiologiska effekter av fytokemikalier. Avhandlingen framhäver potentialen i livsmedelsbearbetning, såsom surdjursjäsning, för att skapa funktionella livsmedel med ökad mängd hälsofrämjande föreningar, och vikten av tarmmikrobiot att mediera hälsoeffekterna.
ACKNOWLEDGEMENTS

The present study was carried out in the Institute of Public Health and Clinical Nutrition, Faculty of Health Sciences, Kuopio campus, University of Eastern Finland, during years 2014–2018.

I am sincerely grateful to my supervisors, Adjunct Professor Kati Hanhineva and Dr Anna-Marja (Annukka) Aura. Kati introduced me to the field of metabolomics, and it was easy to get excited about the research following her enthusiasm. Not only did Kati offer me great and friendly supervision but also an opportunity to travel around Europe and as far as Japan to promote our work and participate in collaborative projects. Annukka’s supervision was indispensable in the *in vitro* experiments, which are a key part of this Doctoral study. It was also a pleasure to get to know such warm personality, perhaps unusual among us Finns. I think the years spent for the doctoral studies were the years best spent in my life so far, mainly thanks to my supervisors.

I would like to thank all the colleagues and collaborators who worked with the study and made it possible. Professor Kaisa Poutanen was involved in the study right from the beginning. Her renowned expertise in cereal research was very valuable for this study and I appreciated her straightforward feedback. Assistant Professor Kati Katina has long experience in the study and development of sourdough. Her critical comments regarding this aspect of the study were extremely important. In VTT, Dr Emilia Nordlund, Dr Outi Mattila, Dr Natalia Rosa-Sibakov and Dr Ismo Mattila were the experts in the cereal bioprocessing and baking technology. With their contribution, we were able to acquire most of our study samples and relate our findings to their previous work. For Study IV, I relied heavily on the work performed by our collaborators at Wageningen University & Research, Dr Klaudyna Borewicz and Professor Hauke Smidt, who performed the microbial community analysis and helped with the challenging manuscript revision. Colleagues from our research group, Dr Olli Kärkkäinen, Dr Jenna Jokkala, and Dr Iman Zarei, also contributed to the study and importantly have also provided me peer support during my studies. The other co-authors equally acknowledged are Professor Seppo Auriola (School of Pharmacy, UEF) and Professor Valérie Micard (SupAgro–INRA–University of Montpellier–CI-RAD). There are also many amazing colleagues that have not been directly involved in this work but whom I have had a pleasure to meet during the years in the 4th floor of Mediteknia building, elsewhere in the UEF campus, in the COST POSITIVe meetings, various conferences, and research visits. I hope the future will bring new interesting collaborations for us.

This PhD work would have not been feasible without financial support from several sources. I gratefully acknowledge the Faculty of Health Sciences (University of Eastern Finland) and Lantmännen Research Foundation for directly funding my research work, Academy of Finland for providing funding via research projects, Euro-
pean Cooperation in Science and Technology (COST) for funding travels to researcher network meetings across Europe and two short-term scientific missions to Lisbon, and Lipoquality project (funded by Japan Society for the Promotion of Science) for granting collaborative research visits to RIKEN Institute in Yokohama.

Finally, I want to thank my family and friends for their tremendous support in this journey. Although I was in a lucky position to be funded throughout the study, pursuing a PhD is a mountain to climb. Without all these people (and two cats) who have greatly contributed to the happiness in my life, I might have stumbled on the way and couldn’t be where I am today. You know who you are without mentioning all the names here, and you are very dear to me now and always.

Kuopio, April 2019

Ville Koistinen
LIST OF ORIGINAL PUBLICATIONS

This dissertation is based on the following original publications:


The publications were adapted with the permission of the copyright owners.
CONTENTS

ABSTRACT .................................................................................................................. 7
TIIVISTELMÄ ............................................................................................................. 9
ABSTRAKT ................................................................................................................. 11
ACKNOWLEDGEMENTS .......................................................................................... 13
1 INTRODUCTION ...................................................................................................... 21
2 REVIEW OF THE LITERATURE .............................................................................. 23
  2.1 HEALTH IMPLICATIONS OF WHOLE GRAINS .............................................. 23
    2.1.1 Proposed mechanisms .............................................................................. 24
    2.1.2 The rye factor .......................................................................................... 25
  2.2 PHYTOCHEMICALS IN WHOLE GRAINS ....................................................... 26
    2.2.1 Alkylresorcinols ...................................................................................... 26
    2.2.2 Benzoxazinoids ....................................................................................... 30
    2.2.3 Betaines ................................................................................................... 34
    2.2.4 Carotenoids ............................................................................................. 36
    2.2.5 Flavonoids ............................................................................................... 38
    2.2.6 Lignans .................................................................................................... 44
    2.2.7 Phenolic acids ........................................................................................ 48
    2.2.8 Phytosterols ............................................................................................. 53
    2.2.9 Tocols ....................................................................................................... 57
  2.3 EFFECT OF FOOD PROCESSING ON WHOLE GRAIN PHYTOCHEMICALS .... 58
    2.3.1 Sourdough ............................................................................................... 58
    2.3.2 Enzymatic bioprocessing ....................................................................... 59
  2.4 GUT MICROBIAL METABOLISM OF WHOLE GRAIN PHYTOCHEMICALS .... 60
    2.4.1 Microbial metabolism of whole grain phytochemicals ......................... 61
  2.5 MASS SPECTROMETRY-BASED NONTARGETED ANALYSIS OF PHYTOCHEMICALS ........................................................................................................ 63
    2.5.1 Chromatographic separation .................................................................. 64
    2.5.2 Applications of mass spectrometry ......................................................... 65
    2.5.3 Sample preparation .................................................................................. 66
    2.5.4 Data analysis ............................................................................................ 66
    2.5.5 Metabolite identification ........................................................................ 68
3 AIMS OF THE STUDY .............................................................................................. 70
4 GENERAL EXPERIMENTAL PROCEDURES ......................................................... 71
5 CHANGES IN THE PHYTOCHEMICAL PROFILE OF RYE BRAN INDUCED BY ENZYMATIC BIOPROCESSING AND SOURDOUGH FERMENTATION (STUDY I) ................................................................. 74
  5.1 Abstract ........................................................................................................... 74
  5.2 Introduction .................................................................................................... 74
  5.3 Materials and methods .................................................................................. 77
  5.4 Results and discussion .................................................................................. 79
  5.5 Conclusions ................................................................................................... 93
  5.6 Acknowledgements ....................................................................................... 93

6 METABOLIC PROFILING OF SOURDOUGH FERMENTED WHEAT AND RYE BREAD (STUDY II) ................................................................. 94
  6.1 Abstract ........................................................................................................... 94
  6.2 Introduction .................................................................................................... 94
  6.3 Materials and methods .................................................................................. 95
  6.4 Results ............................................................................................................ 99
  6.5 Discussion ..................................................................................................... 106
  6.6 Acknowledgements ....................................................................................... 109

7 EFFECT OF BIOPROCESSING ON THE IN VITRO COLONIC MICROBIAL METABOLISM OF PHENOLIC ACIDS FROM RYE BRAN FORTIFIED BREADS (STUDY III) ..................................................... 110
  7.1 Abstract ........................................................................................................... 110
  7.2 Introduction .................................................................................................... 111
  7.3 Materials and methods .................................................................................. 112
  7.4 Results ............................................................................................................ 115
  7.5 Discussion ..................................................................................................... 123
  7.6 Acknowledgements ....................................................................................... 127

8 CONTRIBUTION OF GUT MICROBIOTA TO METABOLISM OF DIETARY GLYCINE BETAIN IN MICE AND IN VITRO COLONIC FERMENTATION (STUDY IV) ................................................................. 128
  8.1 Abstract ........................................................................................................... 128
  8.2 Background .................................................................................................... 129
  8.3 Materials and methods .................................................................................. 130
  8.4 Results ............................................................................................................ 134
  8.5 Discussion ..................................................................................................... 140
  8.6 Conclusions ................................................................................................... 143
  8.7 Declarations .................................................................................................. 143

9 SYNOPSIS OF THE RESULTS ........................................................................... 145

10 GENERAL DISCUSSION ................................................................................. 147
  10.1 Utilisation and relevance of the results ......................................................... 147
  10.2 Strengths and limitations of the methodology ............................................. 148

11 CONCLUSIONS ................................................................................................. 152

REFERENCES ....................................................................................................... 154

APPENDICES ....................................................................................................... 180
## ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>APCI</td>
<td>atmospheric pressure chemical ionisation</td>
</tr>
<tr>
<td>APPI</td>
<td>atmospheric pressure photo-ionization</td>
</tr>
<tr>
<td>BCAA</td>
<td>branched-chain amino acid (isoleucine, leucine, valine)</td>
</tr>
<tr>
<td>CE</td>
<td>capillary electrophoresis</td>
</tr>
<tr>
<td>CNS</td>
<td>central nervous system</td>
</tr>
<tr>
<td>EI</td>
<td>electron ionisation</td>
</tr>
<tr>
<td>ER</td>
<td>oestrogen receptor</td>
</tr>
<tr>
<td>ESI</td>
<td>electrospray ionization</td>
</tr>
<tr>
<td>FODMAP</td>
<td>fermentable oligosaccharides, disaccharides, monosaccharides, and polyols</td>
</tr>
<tr>
<td>GC–MS</td>
<td>gas chromatography–mass spectrometry</td>
</tr>
<tr>
<td>HILIC</td>
<td>hydrophilic interaction liquid chromatography</td>
</tr>
<tr>
<td>HPLC</td>
<td>high-performance liquid chromatography</td>
</tr>
<tr>
<td>IL</td>
<td>interleukin, a group of cytokines working as immunomodulators</td>
</tr>
<tr>
<td>IT</td>
<td>ion trap detector in MS</td>
</tr>
<tr>
<td>LC–MS</td>
<td>liquid chromatography–mass spectrometry</td>
</tr>
<tr>
<td>LDL</td>
<td>low-density lipoprotein</td>
</tr>
<tr>
<td>log $P$</td>
<td>logarithm of the octanol–water partition coefficient (measure of lipophilicity)</td>
</tr>
<tr>
<td>LTQ</td>
<td>linear trap quadrupole MS</td>
</tr>
<tr>
<td>MS</td>
<td>mass spectrometry</td>
</tr>
<tr>
<td>$m/z$</td>
<td>mass-to-charge ratio of an ion ($m$ corresponds to unified atomic mass when $z = 1$)</td>
</tr>
<tr>
<td>NMR</td>
<td>nuclear magnetic resonance (spectroscopy)</td>
</tr>
<tr>
<td>PUFA</td>
<td>polyunsaturated fatty acid (with 2 or more double bonds in the hydrocarbon chain)</td>
</tr>
<tr>
<td>QTOF-MS</td>
<td>quadrupole time-of-flight mass spectrometry</td>
</tr>
<tr>
<td>RP</td>
<td>reversed-phase (liquid chromatography)</td>
</tr>
<tr>
<td>SCFA</td>
<td>short-chain fatty acid (containing ≤ 5 carbons)</td>
</tr>
<tr>
<td>SDG</td>
<td>secoisolariciresinol diglucoside, a lignan compound</td>
</tr>
<tr>
<td>TNF</td>
<td>tumour necrosis factor, a protein family working as immunomodulators</td>
</tr>
<tr>
<td>UHPLC</td>
<td>ultra-high-performance liquid chromatography; same as UPLC</td>
</tr>
</tbody>
</table>
1 INTRODUCTION

Cereal grains are edible dry fruits, i.e. caryopses, produced by certain plant species belonging to the monocot family Poaceae (the grasses). The most important such species, in order of the worldwide production of their grains in 2016, are maize (Zea mays L.), wheat (Triticum sp.), rice (Oryza sativa L. and Oryza glaberrima Steud.), barley (Hordeum vulgare L.), millet (Panicum miliaceum L.), oats (Avena sativa L.), triticale (× Triticosecale), and rye (Secale cereale L.) [1]. Besides the cereals, several grains from other plant families, including amaranth and quinoa from Amaranthaceae and buckwheat from Polygonaceae, are used in a similar manner. These species are referred to as pseudocereals. Cereals are used as staple food in many different forms, one of the most important being bread, usually produced by baking from a dough containing flour, water and a leavening agent, such as yeast and lactic acid bacteria. This thesis will focus on common wheat (Triticum aestivum L.) and rye, the main cereals used for bread production in Northern Europe.

Common wheat accounts for 95% of worldwide wheat production and is one of the main sources of energy (up to 20%) in humans [2]. It putatively originates from a narrow region stretching from Armenia to Caspian Sea, where it emerged as a hybrid between wild emmer wheat (Triticum turgidum ssp. dicoccoides) and another grass species, Aegilops tauschii. Rye, on the other hand, is widely used in sourdough bread baking particularly in Northern and Eastern Europe, where it is one of the main sources of dietary fibre. Similarly to wheat, it likely has its origins in Asia Minor, in the Mount Ararat and Lake Van region of Eastern Turkey, where it first appeared as a weed in the crops of early domesticated cereals; being more hardy than wheat, it gradually became cultivated especially in harsh climates [3].

Whole grain can be defined as the caryopsis of a cereal grain, either in its intact form or e.g. ground into flour, where all the components of the original grain – the endosperm, germ and bran – remain in the same relative proportions [4]. However, there is no universally accepted definition of what can be regarded as a whole-grain product, since depending on the country, the actual whole grain content of a product labelled as whole-grain may be as low as 50% on a dry weight basis, such as in Sweden and Denmark [5]. Thus, when conducting studies involving whole grains or whole-grain products, it is necessary to elaborate the exact composition of the studied food.

Phytochemicals are secondary metabolites produced by plants. In food science and clinical nutrition, the term refers to compounds originating from plants that are not acting as nutrients but that may exert biological activity in vivo after consumption. The bran section of whole grains is a major source of phytochemicals, such as alkylresorcinols, lignans, and phenolic acids [6, 7]. Before their absorption in the gut, phytochemicals may be metabolised by food processing, endogenous enzymes or colonic microbiota [8]. The conversions occurring in the colon can be studied in vitro using a gastrointestinal model [9].
Studying the overall phytochemical composition of cereal products or other food items is generally carried out utilizing a methodology termed metabolomics. It is the study of the metabolome, i.e. all the metabolites produced by a cell, tissue or a complete living organism [10]. While metabolite profiling and targeted metabolite analysis are applied to measure a predetermined set of compounds from a sample, metabolomics (being non-targeted by nature) aims for a more holistic approach by ideally detecting all the metabolites in a global profiling effort to distinguish the metabolites that are differential between two or more groups of samples, differing in e.g. treatment or origin. The benefit of nontargeted metabolomics is to reveal unknown metabolic processes and biomarkers, creating new hypotheses for further targeted studies.

Mass spectrometry (MS) and nuclear magnetic resonance (NMR) spectroscopy are the two prevailing analytical methods in metabolomics [11, 12]. NMR provides easy quantification, high reproducibility, and structural information; however, MS is the most widely used method in nontargeted plant metabolomics because of its high sensitivity and capability to detect and identify up to thousands of metabolites in a single sample. To facilitate the detection process, MS is nearly always coupled with a chromatographic method, most commonly liquid chromatography (LC) or gas chromatography (GC). In the current study, both LC–MS and GC–MS have been utilised to ensure a wide coverage of metabolites detected and identified.

As discussed in this doctoral thesis, increasing evidence suggests that whole grains are beneficial to health by reducing the risk of several chronic diseases and mortality. However, the mechanisms leading to the positive health outcomes are for the most part unknown and are likely related to the array of phytochemicals present within the cereal bran. The purpose of the study presented in this thesis is to elucidate the way food processing and microbial fermentation affects the metabolic profile and phytochemical content of whole-grain wheat and rye products, and in contrast, how whole-grain diet influences the gut microbial composition. Further, the thesis aims to discuss the implications of the presented results, related to some specific groups of phytochemicals, on previously observed health effects.
2 REVIEW OF THE LITERATURE

2.1 HEALTH IMPLICATIONS OF WHOLE GRAINS

Accumulating epidemiological evidence indicates that whole grain consumption has a protective effect against several non-communicable diseases and all-cause mortality [13-17]. The effect does not disappear after the adjustment for other known risk factors of the diseases. Several studies have found that a high intake of whole grains is associated with a lower risk of coronary heart disease [18, 19], ischemic stroke [20, 21], cardiovascular disease [22], type 2 diabetes mellitus [23-25], and colorectal cancer [26], although some publications have shown unclear associations [27]. In addition to these, a recent meta-analysis by Aune et al. (2016) showed that the consumption of whole grains had a non-linear association with the reduced mortality risk of overall cancer, respiratory disease, and infectious diseases [28]. Among different food groups, only the consumption of whole grains, legumes and fish induced a nearly linear dose-dependent risk reduction of all-cause mortality in another meta-analysis by Schwingshackl et al. (2017) [14], suggesting that a high intake of those foods should be promoted. In contrast, no association with the mortality risk was observed in refined grains.

Several meta-analyses have been published that include cohort studies on the association of whole-grain intake and disease mortality. Aune et al. (2016) [28] reported a 21% decrease in coronary heart disease risk (RR 95% CI 0.73–0.86) related to whole grain intake. For type 2 diabetes, a risk reduction as low as 51% (RR 0.23–1.05) was observed [28], while the meta-analysis by Schwingshackl et al. resulted in a more conservative estimate of 23% (RR 0.71–0.84). The risk reduction on cardiovascular disease was 17% (RR 0.79–0.86) in a meta-analysis by Zhang et al. [29]. For overall cancer mortality, the risk reduction has been reported between 6% (RR 0.87–1.01) [29] and 18% (RR 0.69–0.96) [13]. The evidence seems stronger for the prevention of colorectal cancer, where both Aune et al. and Vieira et al. (2017) reported a 17% reduction in mortality risk (RR 0.78–0.89 and 0.79–0.89, respectively) [26, 30]. In a cohort study by Kyrø et al. (2013) in Scandinavian population, a strong association (45% risk reduction) was found between total whole-grain intake and cancer in the proximal colon (RR 0.30–0.99) but not for women or other types of colorectal cancer [31].

Whole grains may also protect against obesity, which in itself is a major risk factor for several lifestyle diseases. In a study by Liu at al. (2003), middle-aged American women in the highest quintile of dietary fibre (from all sources including whole grains) intake had a 49% lower risk of a major weight gain compared to the lowest quintile, and those women who consumed more whole grains weighed consistently less than those with lower whole grain consumption [32]. Similar results have been observed with men: the consumption of whole-grain breakfast cereals was associated with BMI ≤ 25 (a normal body mass index) [33].
2.1.1 Proposed mechanisms

Currently, the exact mechanisms behind the protective effect of whole grains against disease are far from being deciphered, and many of them are based on associations observed in prospective cohort studies. One of the main obstacles in resolving the mechanisms is the large number of various compounds and fibre components present in cereals, many of which possess individual potential health benefits, and the resulting complexity of the biological effects related to whole-grain intake [34]. However, several mechanisms of action have been proposed, briefly reviewed here.

Some of the beneficial health implications of whole foods have been linked with a synergistic and additive effect of the various phytochemicals present in the foods [35]. This would be consistent with the observations that these compounds do not produce as clear positive health effects when they have been ingested individually as supplements. In cancer prevention, the likely mechanism of action of phytochemicals is the antioxidative effect against the free radical formation and oxidative stress, which can lead to potential cancer-inducing damages in the DNA [36]. The possible effects are not restricted to antioxidative activity but may have other complementary mechanisms, such as the beneficial modulation of cell differentiation and proliferation as well as gene expression, and the induction of apoptosis and enzymatic activity. In the risk reduction of cardiovascular disease, it has been hypothesised that phytochemicals prevent the oxidation of low-density lipoproteins (LDL), which would otherwise promote the development of atherosclerosis [35]. Other possible mechanisms include the reduction of platelet formation and blood pressure as well as the inhibition of the synthesis and absorption of cholesterol. The pathogenic factors leading to the onset of type 2 diabetes include oxidative stress and low-grade inflammation; therefore, the antioxidative and anti-inflammatory properties of phytochemicals may contribute to the protective effect of whole grains against the disease [37].

Epidemiological studies provided first evidence in the 1970s that dietary fibre protects against colorectal cancer, based on the incidence of the disease among populations that had adopted either a low- or high-fibre diet [38]. As whole grains are one of the main sources of dietary fibre in many diets, it is the likely reason why similar protective effect has been observed with whole grains. Several mechanisms have thus far been proposed as possible explanations [39]. Firstly, the increased faecal bulk resulting from fibre intake can dilute potential carcinogens in the gut and also reduce the time they are able to be in contact with the gut epithelial cells. The absorption of complex sugars is delayed, which reduces the postprandial insulin peak. Moreover, the gut microbiota ferments dietary fibre [39] and the phytochemicals within the fibre matrix [40], inhibiting the conversion of primary bile acids into more toxic secondary bile acids, producing short-chain fatty acids, which can reduce cell proliferation and induce apoptosis, and transforming phenolic compounds into microbial metabolites, exerting local antioxidative effects in the lumen. The role of diet in modulating the composition of gut microbiota and the impact of the changes on human health has become increasingly evident in the past two decades of research, and dietary fibre and phytochemicals within whole grains seem to have the capability of modifying
the microbial community in a beneficial way [41-45]. The gut microbial taxa with an observed increase after whole-grain diets include *Akkermansia*, *Bifidobacterium*, *Clostridium*, family *Coriobacteriaceae*, *Dialister*, *Eubacterium*, *Lactobacillus*, *Parasutterella*, *Roseburia*, and *Ruminococcus* [43, 44].

Some of the health effects of whole grains may be explained by the increased satiety caused by whole-grain foods when compared to refined wheat containing the same amount of calories [46-48]. The possible mechanisms leading to lowered feeling of hunger include physical factors, such as the contribution of dietary fibre in the viscosity and encapsulation of the food, which delays the availability of nutrients and changes the kinetics of food digestion. On a molecular level, the ingestion of whole-grain rye bread has been shown to increase the levels of certain plasma metabolites, including ribitol, ribonic acid and indole-3-acetic acid, all of which are related to the metabolism of tryptophan and serotonin, the latter known to decrease hunger [49].

### 2.1.2 The rye factor

Leinonen et al. showed in 1999 that whole-grain rye bread induces a lower postprandial insulin response compared to white wheat bread without differences in the glucose response, indicating that less insulin is required to regulate the glucose levels in plasma after the rye bread intake [50], a phenomenon later coined as the rye factor. A similar effect was observed even with refined (endosperm) rye products, but interestingly, not with white wheat bread with added rye bran [51]. As an indication of the effect being specific to rye instead of whole-grain, Kallio et al. observed reduced inflammatory biomarkers after rye bread and pasta intake compared to whole-grain wheat bread [52]. In a randomized cross-over trial by Roager et al., rye intake was associated with reduced levels of IL-6 [53]. Foods with a low glycaemic index (GI), causing lowered acute insulinaemia, have been demonstrated to lower the risk of type 2 diabetes and cardiovascular disease. The mechanism behind the rye factor is not known but several hypotheses exist. The firmer structure of rye grain compared to wheat could lead to obstruction of amylolysis in the mastication phase and slower the rate of glucose release from the food [54]. Compared to white wheat, whole-grain rye also increases the postprandial plasma concentration of short-chain fatty acids (SCFAs), which can beneficially affect the glucose metabolism [55]. Phytochemicals and certain amino acids, such as branched-chain amino acids (BCAAs) and phenylalanine, are released both in the sourdough baking process typical of rye bread [56] and in the mastication of rye bread [57], and may contribute to insulin metabolism. The fermentation of rye bread with yeast or lactic acid bacteria also alters the texture and structure of the bread, which has been linked with the insulin response in a comparison to unfermented rye crispbread [48]. Overall, it seems that there is more than one underlying mechanism behind the rye factor [46]. To further elucidate whether the phenomenon is indeed specific to rye, more studies including whole-grain wheat (or other species) as a reference instead of only white wheat are warranted.
2.2 PHYTOCHEMICALS IN WHOLE GRAINS

When studying the health effects of specific phytochemical classes present in whole foods, it is challenging to relate the observed effects to certain compounds due to the presence of all the other phytochemicals working in synergy. On the other hand, administering these compounds alone might not result in similar effects. The amounts of compounds present in foods and their bioavailability also needs to be considered when associating health effects with phytochemicals, because for exerting any physiological effect, the compounds must be present in the circulation in a sufficient concentration.

In the below chapters, each phytochemical class commonly found in whole grains is introduced with description of their chemistry, potential bioactivity, and analytical characteristics of the compounds in mass spectrometry.

2.2.1 Alkylresorcinols

Alkylresorcinols (AR) are phenolic lipids with an alkyl chain attached to a resorcinol (1,3-dihydroxybenzene) structure (Figure 2.1). In cereals, the alkyl chain is usually odd-numbered and has 15 to 27 carbons, may be unsaturated or have one to three double bonds, and may include one carbonyl or hydroxyl group in the 2’ or 4’ position (Table 2.1). In alkadiyldiresorcinols, two resorcinol moieties are located at each end of the alkyl chain. Alkylresorcinols have been detected in barley, rye and wheat [58]; they also exist in the fruits of the family Anacardiaceae, including mango [59]. They are almost exclusively located in the bran section of cereal grains [60] and have been proposed as biomarkers of whole grain intake [61, 62]. More specifically, the ratio of C17:0 to C21:0 alkylresorcinol species is 0.1 in wheat and 1.0 in rye [60], suggesting that the C17:0 homologue could be used in determining whole-grain rye intake. The validation of alkylresorcinols as whole-grain intake biomarkers is underway [63, 64]. Studies using food diaries to correlate whole-grain wheat and rye intake with plasma AR concentrations have reported moderate correlation coefficients between 0.32 and 0.52 [64]. The correlation of plasma alkylresorcinol levels with two web-based dietary assessment tools was found statistically significant but weak by Nybacka et al. [65], which may result from limitations in either the reporting tools or in the suitability of ARs as objective intake biomarkers. Although it currently seems that the plasma alkylresorcinol content does not highly correlate with the whole-grain intake, it can be used to classify study subjects into quartiles of WG intake [64].

Biological activity. The observed in vitro effects of alkylresorcinols include antioxidative properties, suppression of colon cancer cell growth, and inhibition of lipoygenase activity, LDL oxidation, \( \gamma \)-tocopherol metabolism and lipolysis in adipose tissue [63]. However, the in vivo results are scarcer and include either animal experiments or indirect results from human intervention studies. These observations include decreased hepatic lipid and increased \( \gamma \)-tocopherol concentrations in rats [66] and an inverse association between alkylresorcinol intake and the levels of plasma non-esterified fatty acids [67]. Recently, alkylresorcinols were shown to protect
against oxidative stress in mice fed with a high-fat diet [68]. The compounds also reduced the plasma cholesterol levels in mice by inhibiting the absorption of cholesterol via the reduction of micellar solubility and promotion of cholesterol excretion [69]. In *Drosophila melanogaster*, alkylresorcinols were shown to activate SIRT1, a sirtuin enzyme that catalyses protein deacetylation, resulting in sirtuin-dependent extended lifespan of the fruit flies [70].

**Analysis.** Due to their relatively high log *P* value of 8.5–13.4, alkylresorcinols are practically insoluble in water [71]. Most organic solvents are suitable for the extraction of alkylresorcinols, in particular acetone and ethyl acetate; methanol can also be used, but it may discriminate against the most lipophilic species, such as C23:0 and C25:0. Several targeted methods exist for the analysis of alkylresorcinols in cereal matrix. Ross et al. developed a GC–MS method for the detection and quantification of alkylresorcinols using ethyl acetate as the extraction solvent [72]. In LC–MS, alkylresorcinols are detectable in both positive and negative ionization mode. They are, however, not as easily quantitated with LC–MS compared to GC–MS, and the process may require additional filtering and purification steps using e.g. solid-phase extraction [71]. Nevertheless, methods coupling HPLC with coulometric array, ultraviolet or fluorescence detection, have been developed [73]. Several nontargeted methods have detected alkylresorcinols from cereal grains and their products [74, 75]; however, in a multiplatform coverage test performed on 13 different platforms deploying nontargeted mass spectrometry, we found that the majority of the unoptimized LC–MS methods were unable to detect the C17:0 and C25:0 species [76].

The chromatographic and mass spectral characteristics of alkylresorcinols include high retention time (in reversed-phase LC) and typical MS/MS fragmentation patterns depending on the saturation of the alkyl chain [58, 77]. In the positive ionization mode, saturated alkylresorcinols form fragment ions at *m/z* 85 and 111, explained by heterolytic α-cleavage of the aromatic group [58, 59], and in the negative mode, typical fragments include *m/z* 81, 122 and 135 [59, 74]. A fragment at *m/z* 124, resulting from McLafferty rearrangement, is commonly seen in GC–MS analyses but not in LC–MS [59, 72], which is possibly attributable to the electron ionization (EI) technique commonly used in gas chromatography. Mono- and polyunsaturated alkylresorcinols typically produce fragments at *m/z* 123, 137, 163, and 177 in the positive mode and at *m/z* 81, 83, 122, and 135 in the negative mode [59]. In hydroxylated alkylresorcinols, a diagnostic characteristic in the positive mode is the neutral loss of water (18 Da) from fragments with high mass. A neutral loss of 42 Da from the molecular ion, putatively assigned as an acyl moiety originating from the phenolic ring [74], can be observed in all alkylresorcinols.
Figure 2.1. Chemical structures of alkylresorcinols and their metabolites detected in (or after the consumption of) whole grains. Two example structures are given for alkylresorcinols possessing a carbonyl or hydroxyl group and unsaturated aliphatic chain.

Table 2.1. Mass spectral characteristics of alkylresorcinols and their metabolites. The origin signifies whether the compound is present in intact whole grain (WG) or as a microbial / endogenous metabolite. MM = monoisotopic neutral molecular mass in Daltons.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Origin</th>
<th>Mol. formula</th>
<th>MM</th>
<th>Fragments m/z (−)</th>
<th>Fragments m/z (+)</th>
<th>Detection method</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3,5-DHBA</td>
<td>microbiota</td>
<td>C₇H₆O₄</td>
<td>154.0266</td>
<td>163, 145, 137, 135, 121, 117, 109</td>
<td></td>
<td>GC–MS</td>
<td>[78]</td>
</tr>
<tr>
<td>3,5-DHPPA</td>
<td>microbiota</td>
<td>C₉H₁₀O₄</td>
<td>182.0579</td>
<td>153, 109, 65, 67, 232</td>
<td></td>
<td>UPLC–QTOF</td>
<td>[79]</td>
</tr>
<tr>
<td>3,5-DHBA sulfate</td>
<td>liver</td>
<td>C₇H₈O₂S</td>
<td>233.9834</td>
<td>81, 83, 122, 277, 308</td>
<td>111, 99, 123, 110</td>
<td>UPLC–QTOF</td>
<td>[80]</td>
</tr>
<tr>
<td>pentadecenylresorcinol (C15:1)</td>
<td>WG</td>
<td>C₁₂H₂₀O₂</td>
<td>318.2559</td>
<td>123</td>
<td></td>
<td>UPLC–TQ</td>
<td>[59]</td>
</tr>
<tr>
<td>pentadecylresorcinol (C15:0)</td>
<td>WG</td>
<td>C₁₂H₁₈O₂</td>
<td>320.2715</td>
<td>123</td>
<td></td>
<td>HPLC–APCI, UPLC–QTOF</td>
<td>[58], [74]</td>
</tr>
<tr>
<td>heptadecatrienylresorcinol (C17:3)</td>
<td>WG</td>
<td>C₁₃H₂₄O₂</td>
<td>342.2559</td>
<td>123, 163, 177, 137, 109, 191</td>
<td></td>
<td>HPLC–APCI</td>
<td>[81]</td>
</tr>
<tr>
<td>heptadecadienylresorcinol (C17:2)</td>
<td>WG</td>
<td>C₁₃H₂₆O₂</td>
<td>344.2715</td>
<td>123, 163, 149, 177, 301</td>
<td></td>
<td>HPLC–APCI, UPLC–QTOF</td>
<td>[58], [74]</td>
</tr>
<tr>
<td>heptadecenylresorcinol (C17:1)</td>
<td>WG</td>
<td>C₁₃H₂₆O₂</td>
<td>346.2872</td>
<td>123</td>
<td></td>
<td>HPLC–APCI, UPLC–QTOF</td>
<td>[58], [74]</td>
</tr>
<tr>
<td>heptadecylresorcinol (C17:0)</td>
<td>WG</td>
<td>C₁₃H₂₈O₂</td>
<td>348.3028</td>
<td>123</td>
<td></td>
<td>HPLC–APCI, UPLC–QTOF</td>
<td>[58], [74]</td>
</tr>
<tr>
<td>C17:1, hydroxy</td>
<td>WG</td>
<td>C₁₃H₂₈O₃</td>
<td>362.2821</td>
<td>123, 345, 163, 125, 149</td>
<td></td>
<td>HPLC–APCI</td>
<td>[58]</td>
</tr>
<tr>
<td>Compound</td>
<td>Origin</td>
<td>Mol. formula</td>
<td>MM</td>
<td>Fragments m/z (−)</td>
<td>Fragments m/z (+)</td>
<td>Detection method</td>
<td>Ref.</td>
</tr>
<tr>
<td>---------------------------</td>
<td>--------</td>
<td>--------------</td>
<td>--------</td>
<td>-------------------</td>
<td>-------------------</td>
<td>-----------------</td>
<td>------</td>
</tr>
<tr>
<td>octadecylenoscorinol (C18:0)</td>
<td>WG</td>
<td>C_{29}H_{46}O_{2}</td>
<td>362.3185</td>
<td>111, 343</td>
<td></td>
<td>HPLC-APCI</td>
<td>[58]</td>
</tr>
<tr>
<td>nonadecatrienoylresorinol (C19:3)</td>
<td>WG</td>
<td>C_{29}H_{46}O_{2}</td>
<td>370.2872</td>
<td>59, 81, 122, 135, 149</td>
<td>123, 163, 137, 177</td>
<td>HPLC-APCI, UPLC-QTOF</td>
<td>[58, 74]</td>
</tr>
<tr>
<td>nonadecadienoylresorinol (C19:2)</td>
<td>WG</td>
<td>C_{29}H_{46}O_{2}</td>
<td>372.3028</td>
<td>81, 122, 135, 149, 191, 329</td>
<td>123, 137, 163, 251</td>
<td>HPLC-APCI, UPLC-QTOF</td>
<td>[58, 74]</td>
</tr>
<tr>
<td>nonadecenoylresorinol (C19:1)</td>
<td>WG</td>
<td>C_{29}H_{46}O_{2}</td>
<td>374.3185</td>
<td>81, 122, 123, 135, 149, 331</td>
<td>123, 248, 197, 124</td>
<td>HPLC-APCI, UPLC-QTOF</td>
<td>[58, 74]</td>
</tr>
<tr>
<td>nonadecylresorinol (C19:0)</td>
<td>WG</td>
<td>C_{29}H_{46}O_{2}</td>
<td>376.3341</td>
<td>81, 122, 333</td>
<td>111</td>
<td>HPLC-APCI, UPLC-QTOF</td>
<td>[58, 74]</td>
</tr>
<tr>
<td>C19:1, hydroxy</td>
<td>WG</td>
<td>C_{29}H_{46}O_{3}</td>
<td>390.3134</td>
<td>123, 149, 163, 125, 373</td>
<td></td>
<td>HPLC-APCI</td>
<td>[58]</td>
</tr>
<tr>
<td>C19:0, oxo</td>
<td>WG</td>
<td>C_{29}H_{46}O_{3}</td>
<td>390.3134</td>
<td>43, 57, 123, 267, 124</td>
<td></td>
<td>HPLC-MS</td>
<td>[82]</td>
</tr>
<tr>
<td>heneicosatrienoylresorinol (C21:3)</td>
<td>WG</td>
<td>C_{31}H_{50}O_{2}</td>
<td>398.3185</td>
<td>81, 122, 135</td>
<td>123, 163, 137, 329</td>
<td>HPLC-APCI, UPLC-QTOF</td>
<td>[58, 74]</td>
</tr>
<tr>
<td>heneicosadienoylresorinol (C21:2)</td>
<td>WG</td>
<td>C_{31}H_{50}O_{2}</td>
<td>400.3341</td>
<td>81, 122, 123, 135, 136, 149, 357</td>
<td>123, 137, 289, 177</td>
<td>HPLC-APCI, UPLC-QTOF</td>
<td>[58, 74]</td>
</tr>
<tr>
<td>heneicosenoylresorinol (C21:1)</td>
<td>WG</td>
<td>C_{31}H_{50}O_{2}</td>
<td>402.3498</td>
<td>81, 122, 135, 149, 191, 359</td>
<td>123, 163, 124, 137</td>
<td>HPLC-APCI, UPLC-QTOF</td>
<td>[58, 74]</td>
</tr>
<tr>
<td>heneicosylresorinol (C21:0)</td>
<td>WG</td>
<td>C_{31}H_{50}O_{2}</td>
<td>404.3654</td>
<td>81, 122, 149, 361</td>
<td>123</td>
<td>HPLC-APCI, UPLC-QTOF</td>
<td>[58, 74]</td>
</tr>
<tr>
<td>C21:1, oxo</td>
<td>WG</td>
<td>C_{31}H_{50}O_{3}</td>
<td>416.3290</td>
<td>124, 267, 123, 293, 166, 137, 294</td>
<td></td>
<td>HPLC-EI/MS</td>
<td>[83]</td>
</tr>
<tr>
<td>C21:1, hydroxy</td>
<td>WG</td>
<td>C_{31}H_{50}O_{3}</td>
<td>418.3447</td>
<td>123, 149, 163, 401</td>
<td></td>
<td>HPLC-APCI</td>
<td>[58]</td>
</tr>
<tr>
<td>C21:0, oxo</td>
<td>WG</td>
<td>C_{31}H_{50}O_{3}</td>
<td>418.3450</td>
<td>295, 124, 123, 321, 166, 137</td>
<td></td>
<td>HPLC-EI/MS</td>
<td>[83]</td>
</tr>
<tr>
<td>tricosadienoylresorinol (C23:2)</td>
<td>WG</td>
<td>C_{33}H_{54}O_{2}</td>
<td>428.3654</td>
<td>71, 81, 122, 135, 159, 177, 337, 385</td>
<td>123, 137, 163, 317</td>
<td>HPLC-APCI, UPLC-QTOF</td>
<td>[58, 74]</td>
</tr>
<tr>
<td>tricosenoylresorinol (C23:1)</td>
<td>WG</td>
<td>C_{33}H_{54}O_{2}</td>
<td>430.3811</td>
<td>81, 122, 330, 387</td>
<td>123, 144, 149, 124</td>
<td>HPLC-APCI, UPLC-QTOF</td>
<td>[58, 74]</td>
</tr>
<tr>
<td>tricosylresorinol (C23:0)</td>
<td>WG</td>
<td>C_{33}H_{54}O_{2}</td>
<td>432.3967</td>
<td>79, 81, 122, 389</td>
<td>123, 391, 199, 141</td>
<td>HPLC-APCI, UPLC-QTOF</td>
<td>[58, 74]</td>
</tr>
<tr>
<td>C23:1, oxo</td>
<td>WG</td>
<td>C_{33}H_{54}O_{3}</td>
<td>444.3603</td>
<td>55, 321, 43, 123, 124</td>
<td></td>
<td>HPLC-MS</td>
<td>[82]</td>
</tr>
<tr>
<td>C23:1, hydroxy</td>
<td>WG</td>
<td>C_{33}H_{54}O_{3}</td>
<td>446.3760</td>
<td>149, 163, 429, 123</td>
<td></td>
<td>HPLC-APCI</td>
<td>[58]</td>
</tr>
<tr>
<td>C23:0, oxo</td>
<td>WG</td>
<td>C_{33}H_{54}O_{3}</td>
<td>446.3760</td>
<td>43, 57, 123, 124, 323</td>
<td></td>
<td>HPLC-MS</td>
<td>[82]</td>
</tr>
</tbody>
</table>
2.2.2 Benzoazinoids

Benzoazinoids (BX) are aromatic and heterocyclic compounds containing an oxazine group (a 5- or 6-carbon ring containing an oxygen and a nitrogen atom) attached to a benzene ring (Figure 2.2). Based on the more exact structure, they can be further divided into benzoazolinones, lactams, and hydroxamic acids (Table 2.2). Benzoazinoids are produced mainly by Poaceae species and by a few dicot plants [85] and they have been thus far found in the grains of maize, rye and wheat [86, 87]. The total benzoazinoid content in wheat is only 5 µg/g DM but about 50 times higher in rye. Benzoazinoids are essentially defence molecules, stored as glycosides and rapidly hydrolysed into their active aglycone forms when the plant faces damage [88].

**Biological activity.** As reviewed by Andersson et al. and Adhikari et al., benzoazinoids possess a variety of potential health benefits [6, 89]. HBOA and DIBOA have shown cytotoxic effects *in vitro* on cancer cells, such as prostate cancer [90]. Some benzoazinoids may have immunoregulatory properties, such as the inhibition of histamine release by HBOA, HMBOA, and MBOA [91] and the *in vitro* inhibition of nitric oxide, prostaglandin E2 and IL-6 production by APO, a transformation product of benzoazinoids [92]. In contrast, dietary benzoazinoids enhanced the IL-1β, IL-6, and TNF-α responses to bacterial lipopolysaccharide (LPS) in a clinical study setting [93]. An inhibitory effect of certain benzoazinoids against pathogenic bacteria and the reverse-transcriptase activity of human immunodeficiency virus (HIV) has been demonstrated, as reviewed by Adhikari et al. [89]. MBOA has been observed to stimulate the reproductive system of various animals by increasing the synthesis of FSH, resulting in higher pregnancy incidence and testicular hypertrophy [89]. Further-
more, as part of a patent, MBOA has been administered as a supplement to overweight humans, inducing a significant weight reduction (0.9 kg in 30 days) compared to control by promoting lipolysis and decreasing the desire for foods rich in carbohydrates and saturated fats [94]. MBOA has also been tested as a CNS stimulant; it was shown to reduce depression and anxiety as well as improve sexual function [94, 95]. Overall, benzoxazinoids show considerable potential as pharmaceuticals, possibly due to their chemical structure which resembles endogenous human hormones and neurotransmitters, such as melatonin and serotonin. However, the evidence behind the effects is originated mostly from in vitro and animal experiments, and more randomized controlled trials are needed to establish the health benefits of benzoxazinoids.

**Analysis.** As semi-polar metabolites, benzoxazinoids are especially suitable for detection with LC–MS. Typically, they can be extracted from biological samples with 75% or 80% methanol [86, 87, 96]. In LC–MS/MS, benzoxazinoid glycosides easily lose one or several hexose units, producing a neutral loss of 162 or 324 Da in the fragmentation pattern [87]. The aglycone may also lose a water molecule (neutral loss of 18 Da), 1–2 CO moieties (neutral loss of 28 Da each), or a methoxy radical (neutral loss of 15 Da). In the negative ionization mode, \( m/z \) 108 and 118 are characteristic fragment ions for lactams and hydroxamic acids (Table 2.2).
Figure 2.2. Chemical structures of benzoazinoids and their metabolites detected (or after the consumption of) whole grains.

Table 2.2. Mass spectral characteristics of benzoazinoids and their metabolites. The origin signifies whether the compound is present in intact whole grain (WG) or as a microbial/endogenous metabolite. MM = monoisotopic neutral molecular mass in Daltons.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Origin</th>
<th>Mol. formula</th>
<th>MM</th>
<th>Fragments m/z (−)</th>
<th>Fragments m/z (+)</th>
<th>Detection method</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOA</td>
<td>WG</td>
<td>C_8H_8N_2O_2</td>
<td>135.0320</td>
<td>112, 91, 78</td>
<td>136</td>
<td>UPLC–QTOF</td>
<td>[87]</td>
</tr>
<tr>
<td>HPAA</td>
<td>uncertain</td>
<td>C_8H_8N_2O_2</td>
<td>151.0633</td>
<td>93, 94, 107, 108</td>
<td></td>
<td>UPLC–QTOF</td>
<td>[74]</td>
</tr>
<tr>
<td>HBOA</td>
<td>WG</td>
<td>C_8H_8N_2O_2</td>
<td>165.0426</td>
<td>107, 108, 118, 134</td>
<td></td>
<td>UPLC–QTOF</td>
<td>[74]</td>
</tr>
<tr>
<td>MBOA</td>
<td>WG</td>
<td>C_8H_8N_2O_2</td>
<td>165.0426</td>
<td>149</td>
<td></td>
<td>HPLC–QTOF</td>
<td>[97, QTRAP 98]</td>
</tr>
<tr>
<td>HHPAA</td>
<td>uncertain</td>
<td>C_8H_8N_2O_2</td>
<td>167.0582</td>
<td>66, 94, 108, 118</td>
<td></td>
<td>UPLC–QTOF</td>
<td>[74]</td>
</tr>
<tr>
<td>2-aminophenol sulfate</td>
<td>liver</td>
<td>C_6H_8N_2O_4S</td>
<td>189.0096</td>
<td>109, 108, 92</td>
<td></td>
<td>UPLC–QTOF</td>
<td>[80, 99]</td>
</tr>
<tr>
<td>HMBOA</td>
<td>WG</td>
<td>C_8H_8N_2O_3</td>
<td>195.0532</td>
<td>107, 108</td>
<td></td>
<td>UPLC–QTOF</td>
<td>[74]</td>
</tr>
<tr>
<td>DIMBOA</td>
<td>WG</td>
<td>C_8H_8N_2O_3</td>
<td>211.0481</td>
<td>66, 71, 94, 115, 122, 166</td>
<td></td>
<td>UPLC–QTOF</td>
<td>[74]</td>
</tr>
<tr>
<td>APO</td>
<td>uncertain</td>
<td>C_8H_8N_2O_2</td>
<td>212.0586</td>
<td>213, 185</td>
<td></td>
<td>HPLC–TQ</td>
<td>[100]</td>
</tr>
<tr>
<td>HPAA sulfate</td>
<td>liver</td>
<td>C_8H_8N_2O_4S</td>
<td>231.0201</td>
<td>150, 164, 107, 108, 193</td>
<td></td>
<td>UPLC–QTOF</td>
<td>[101]</td>
</tr>
<tr>
<td>Compound</td>
<td>Origin</td>
<td>Mol. formula</td>
<td>MM</td>
<td>Fragments m/z (+)</td>
<td>Fragments m/z (+)</td>
<td>Detection method</td>
<td>Ref.</td>
</tr>
<tr>
<td>------------------</td>
<td>------------------</td>
<td>--------------</td>
<td>-------</td>
<td>-----------------------</td>
<td>-----------------------</td>
<td>------------------</td>
<td>------</td>
</tr>
<tr>
<td>DIM$_2$BOA</td>
<td>WG</td>
<td>C$<em>{15}$H$</em>{15}$N$_{6}$O$_6$</td>
<td>241.0586</td>
<td>66, 71, 78, 94, 106, 115, 122, 124, 196</td>
<td>94, 106, 115, 122, 124, 196</td>
<td>UPLC-QTOF [74]</td>
<td></td>
</tr>
<tr>
<td>AMPO</td>
<td>uncertain</td>
<td>C$<em>{13}$H$</em>{19}$N$_{4}$O$_3$</td>
<td>242.0691</td>
<td>243, 228</td>
<td>243, 228</td>
<td>HPLC–TQ [100]</td>
<td></td>
</tr>
<tr>
<td>HHPAA sulfate</td>
<td>liver</td>
<td>C$<em>8$H$</em>{16}$NO$_3$S</td>
<td>247.0151</td>
<td>166, 108, 118</td>
<td>166, 108, 118</td>
<td>UPLC-QTOF [101]</td>
<td></td>
</tr>
<tr>
<td>AAPO</td>
<td>uncertain</td>
<td>C$<em>{13}$H$</em>{19}$N$_{4}$O$_3$</td>
<td>254.0691</td>
<td>255, 213</td>
<td>255, 213</td>
<td>HPLC–TQ [100]</td>
<td></td>
</tr>
<tr>
<td>DIBOA sulfate</td>
<td>liver</td>
<td>C$<em>{13}$H$</em>{19}$N$_{4}$O$_3$</td>
<td>261.0334</td>
<td>180, 162, 137, 118, 108</td>
<td>180, 162, 137, 118, 108</td>
<td>UPLC-QTOF [101]</td>
<td></td>
</tr>
<tr>
<td>AAMPO</td>
<td>uncertain</td>
<td>C$<em>{13}$H$</em>{19}$N$_{4}$O$_3$</td>
<td>284.0797</td>
<td>285, 243</td>
<td>285, 243</td>
<td>HPLC–TQ [100]</td>
<td></td>
</tr>
<tr>
<td>HBOA hexose</td>
<td>WG</td>
<td>C$<em>{13}$H$</em>{19}$N$_{4}$O$_3$</td>
<td>327.0954</td>
<td>164, 136, 118, 108</td>
<td>164, 136, 118, 108</td>
<td>UPLC-QTOF [87]</td>
<td></td>
</tr>
<tr>
<td>HBOA glucuronide</td>
<td>liver</td>
<td>C$<em>{13}$H$</em>{19}$N$_{4}$O$_3$</td>
<td>341.0747</td>
<td>FT-ICR-MS</td>
<td>FT-ICR-MS</td>
<td></td>
<td>[102]</td>
</tr>
<tr>
<td>DIBOA hexose</td>
<td>WG</td>
<td>C$<em>{13}$H$</em>{19}$N$_{4}$O$_3$</td>
<td>343.0903</td>
<td>180, 162, 134, 118, 108</td>
<td>180, 162, 134, 118, 108</td>
<td>UPLC-QTOF [87]</td>
<td></td>
</tr>
<tr>
<td>HHPAA glucuronide</td>
<td>liver</td>
<td>C$<em>{13}$H$</em>{19}$N$_{4}$O$_3$</td>
<td>343.0903</td>
<td>FT-ICR-MS</td>
<td>FT-ICR-MS</td>
<td></td>
<td>[102]</td>
</tr>
<tr>
<td>DHBOA hexose</td>
<td>WG</td>
<td>C$<em>{13}$H$</em>{19}$N$_{4}$O$_3$</td>
<td>343.0903</td>
<td>FT-ICR-MS</td>
<td>FT-ICR-MS</td>
<td></td>
<td>[102]</td>
</tr>
<tr>
<td>HMBOA hexose</td>
<td>WG</td>
<td>C$<em>{13}$H$</em>{19}$N$_{4}$O$_3$</td>
<td>357.1060</td>
<td>108, 132, 150, 175, 194</td>
<td>108, 132, 150, 175, 194</td>
<td>UPLC-QTOF [74]</td>
<td></td>
</tr>
<tr>
<td>DIBOA dihexose</td>
<td>WG</td>
<td>C$<em>{13}$H$</em>{19}$N$_{4}$O$_3$</td>
<td>373.1009</td>
<td>210</td>
<td>210</td>
<td>UPLC-QTOF [103]</td>
<td></td>
</tr>
<tr>
<td>DIBOA hexose</td>
<td>WG</td>
<td>C$<em>{13}$H$</em>{19}$N$_{4}$O$_3$</td>
<td>373.1009</td>
<td>164</td>
<td>164</td>
<td>UPLC-QTOF [103]</td>
<td></td>
</tr>
<tr>
<td>HBOA dihexose</td>
<td>WG</td>
<td>C$<em>{13}$H$</em>{19}$N$_{4}$O$_3$</td>
<td>489.1482</td>
<td>221, 164, 136, 118, 108</td>
<td>221, 164, 136, 118, 108</td>
<td>UPLC-QTOF [87]</td>
<td></td>
</tr>
<tr>
<td>DIBOA dihexose</td>
<td>WG</td>
<td>C$<em>{13}$H$</em>{19}$N$_{4}$O$_3$</td>
<td>505.1432</td>
<td>506, 344, 325, 182, 164, 136, 80, 78, 59</td>
<td>506, 344, 325, 182, 164, 136, 80, 78, 59</td>
<td>UPLC-QTOF [87]</td>
<td></td>
</tr>
<tr>
<td>DHBOA dihexose</td>
<td>WG</td>
<td>C$<em>{13}$H$</em>{19}$N$_{4}$O$_3$</td>
<td>505.1432</td>
<td>164</td>
<td>164</td>
<td>UPLC-QTOF [98]</td>
<td></td>
</tr>
<tr>
<td>HMBOA dihexose</td>
<td>WG</td>
<td>C$<em>{13}$H$</em>{19}$N$_{4}$O$_3$</td>
<td>519.1588</td>
<td>164</td>
<td>164</td>
<td>UPLC-QTOF [98]</td>
<td></td>
</tr>
<tr>
<td>DIBOA dihexose</td>
<td>WG</td>
<td>C$<em>{13}$H$</em>{19}$N$_{4}$O$_3$</td>
<td>535.1537</td>
<td>210, 192, 177</td>
<td>210, 192, 177</td>
<td>UPLC-QTOF [103]</td>
<td></td>
</tr>
<tr>
<td>DIBOA dihexose</td>
<td>WG</td>
<td>C$<em>{13}$H$</em>{19}$N$_{4}$O$_3$</td>
<td>535.1537</td>
<td>536, 374, 212, 194, 166</td>
<td>536, 374, 212, 194, 166</td>
<td>UPLC-QTOF [87]</td>
<td></td>
</tr>
<tr>
<td>HBOA trihexose</td>
<td>WG</td>
<td>C$<em>{13}$H$</em>{19}$N$_{4}$O$_3$</td>
<td>651.2011</td>
<td>164</td>
<td>164</td>
<td>UPLC-QTOF [103]</td>
<td></td>
</tr>
<tr>
<td>DIBOA trihexose</td>
<td>WG</td>
<td>C$<em>{13}$H$</em>{19}$N$_{4}$O$_3$</td>
<td>667.1960</td>
<td>180, 162</td>
<td>180, 162</td>
<td>UPLC-QTOF [98, 103]</td>
<td></td>
</tr>
<tr>
<td>HMBOA trihexose</td>
<td>WG</td>
<td>C$<em>{13}$H$</em>{19}$N$_{4}$O$_3$</td>
<td>681.2116</td>
<td>194</td>
<td>194</td>
<td>UPLC-QTOF [103]</td>
<td></td>
</tr>
<tr>
<td>DIBOA trihexose</td>
<td>WG</td>
<td>C$<em>{13}$H$</em>{19}$N$_{4}$O$_3$</td>
<td>697.2065</td>
<td>164</td>
<td>164</td>
<td>UPLC-QTOF [103]</td>
<td></td>
</tr>
<tr>
<td>DIBOA tetrahexo</td>
<td>WG</td>
<td>C$<em>{13}$H$</em>{19}$N$_{4}$O$_3$</td>
<td>829.2488</td>
<td>134</td>
<td>134</td>
<td>UPLC-QTOF [103]</td>
<td></td>
</tr>
<tr>
<td>HMBOA tetrahexo</td>
<td>WG</td>
<td>C$<em>{13}$H$</em>{19}$N$_{4}$O$_3$</td>
<td>843.2645</td>
<td>134</td>
<td>134</td>
<td>UPLC-QTOF [103]</td>
<td></td>
</tr>
<tr>
<td>DIBOA tetrahexo</td>
<td>WG</td>
<td>C$<em>{13}$H$</em>{19}$N$_{4}$O$_3$</td>
<td>859.2594</td>
<td>134</td>
<td>134</td>
<td>UPLC-QTOF [103]</td>
<td></td>
</tr>
</tbody>
</table>
2.2.3 Betaines

Betaines, by their extensive definition, are zwitterionic compounds with a cationic onium atom that is not bearing any hydrogens [104]. Originally, the term was limited to glycine betaine \((N,N,N\text{-trimethylglycine})\), which was first discovered from sugar beet \((Beta vulgaris L.)\) in the 19th century, hence giving the name for the compound group. Several other betaines than glycine betaine are derived from amino acids and other small compounds (Figure 2.3), and at least 25 such compounds have been detected in living organisms [105]. Recently, many of these have also been detected in small amounts from cereal flours and bread [56, 106].

Humans obtain betaines from food, mostly as glycine betaine or its precursor, choline [107]. Free choline can be transformed into glycine betaine by a two-step enzymatic process occurring mainly in the mitochondria of liver and kidneys [108], thus making glycine betaine considered as a non-essential nutrient. The main source of dietary betaines in the western diet is wheat, also in its refined form [109]. Other important sources include shellfish, spinach, and sugar beets [108]. Among cereals and pseudocereals, high or moderate amounts of glycine betaine or choline can be found (besides wheat) from amaranth, barley, buckwheat, quinoa and rye, while their levels are relatively low in maize, millet, oats, rice, and sorghum [109, 110]. The average daily intake of glycine betaine has been estimated to range from 0.13 to 2.5 g/day, depending on the diet [108, 109].

**Biological activity.** Many betaines are known to be biologically active: they have an osmoprotective role, protecting the cell from dehydration, osmotic stress, extreme temperatures and high salinity [107]. In addition, they are participating as methyl donors in transmethylation reactions of the methionine cycle, which is in turn occurring in several vital processes, such as the transformation of homocysteine into methionine [108]. Trigonelline, the \(N\)-methylated form of niacin, has been identified as a plant oestrogen [111] and has shown antibacterial (anti-cariogenic) effects [112]. However, the physiological significance of amino acid–derived betaines and their relation to maintaining health is not known [80]. Recently, we have characterized a novel betainized compound, 5-aminovaleric acid betaine (5-AVAB), which is associated with the intake of whole grains and accumulating in metabolically active tissues, such as heart [113]. In the study, we demonstrated that 5-AVAB inhibits the intake of L-carnitine into mouse cardiomyocytes, resulting in decreased β-oxidation of fatty acids and reduced oxygen consumption without compromising mitochondrial respiration, which provides one potential mechanism by which diets rich in whole grains beneficially affect cardiovascular health.

**Analysis.** Betaines do not lend themselves easily for analysis: they are poorly volatile, lack fluorescence and absorb UV weakly at low wavelengths [114]. Because of their inherent positive charge, betaines are suitable for analysis with mass spectrometry; however, they are also highly hydrophilic and therefore challenging for typical reversed-phase chromatography. Therefore, most LC–MS platforms have applied hydrophilic interaction chromatography (HILIC) in betaine analysis [80, 115]. Suita-
ble solvents for the extraction of betaines from plant matrix include water and methanol [114, 116]. In the LC–MS/MS positive mode, betaines with a trimethylammonium group exhibit characteristic fragments at \( m/z \) 58 and 59 (Table 2.3, [117]), while trigonelline, proline betaine, and pipecolic acid betaine produce different characteristic MS/MS spectra (Table 2.3).

![Figure 2.3. Chemical structures of betaines and their metabolites detected in (or after the consumption of) whole grains.](image)

**Table 2.3.** Mass spectral characteristics of betaines in whole grains. The origin signifies whether the compound is present in intact whole grain (WG) or as a microbial/endogenous metabolite. MM = monoisotopic neutral molecular mass in Daltons.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Origin</th>
<th>Mol. formula</th>
<th>MM</th>
<th>Fragments ( m/z ) (+)</th>
<th>Detection method</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>glycine betaine</td>
<td>WG</td>
<td>C(_2)H(_7)NO(_2)</td>
<td>117.0790</td>
<td>58, 59, 118</td>
<td>UPLC–QTOF</td>
<td>[80]</td>
</tr>
<tr>
<td>alanine betaine</td>
<td>WG</td>
<td>C(_3)H(_9)NO(_2)</td>
<td>131.0946</td>
<td>59, 58, 132, 70</td>
<td>UPLC–QTOF</td>
<td>[80]</td>
</tr>
<tr>
<td>trigonelline</td>
<td>WG</td>
<td>C(_5)H(_8)N(_2)O (_2)</td>
<td>137.0477</td>
<td>65, 79, 78, 92, 53, 66, 51, 93, 94</td>
<td>UPLC–QTOF</td>
<td>[80]</td>
</tr>
<tr>
<td>proline betaine (stachydrine)</td>
<td>WG</td>
<td>C(_4)H(_9)NO(_2)</td>
<td>143.0946</td>
<td>144, 58, 84, 102</td>
<td>UPLC–QTOF</td>
<td>[80]</td>
</tr>
<tr>
<td>4-aminobutyric acid betaine</td>
<td>WG</td>
<td>C(_5)H(_9)NO(_2)</td>
<td>145.1103</td>
<td>87, 43, 60, 45, 93, 94</td>
<td>UPLC–QTOF</td>
<td>[105]</td>
</tr>
<tr>
<td>pipecolic acid betaine</td>
<td>WG</td>
<td>C(_6)H(_9)NO(_2)</td>
<td>157.1103</td>
<td>158, 58, 112, 98, 67, 140, 102, 99</td>
<td>UPLC–QTOF</td>
<td>[80]</td>
</tr>
<tr>
<td>Compound</td>
<td>Origin</td>
<td>Mol. formula</td>
<td>MM</td>
<td>Fragments m/z (+)</td>
<td>Detection method</td>
<td>Ref.</td>
</tr>
<tr>
<td>----------------------</td>
<td>--------</td>
<td>--------------</td>
<td>-------</td>
<td>------------------</td>
<td>------------------</td>
<td>------</td>
</tr>
<tr>
<td>valine betaine</td>
<td>WG</td>
<td>C₆H₁₇NO₂</td>
<td>159.1259</td>
<td>60, 59, 58, 160</td>
<td>UPLC–QToF</td>
<td>[56]</td>
</tr>
<tr>
<td>5-AVAB</td>
<td>uncertain</td>
<td>C₆H₁₇NO₂</td>
<td>159.1259</td>
<td>55, 60, 101, 83, 59, 160</td>
<td>UPLC–QToF</td>
<td>[113]</td>
</tr>
<tr>
<td>isoleucine betaine</td>
<td>WG</td>
<td>C₆H₁₇NO₂</td>
<td>173.1416</td>
<td>60, 59, 58, 73</td>
<td>ESI–IT</td>
<td>[105]</td>
</tr>
<tr>
<td>glutamine betaine</td>
<td>WG</td>
<td>C₆H₁₄N₂O₃</td>
<td>188.1161</td>
<td>130, 60, 102, 85, 44, 172</td>
<td>ESI–IT</td>
<td>[105]</td>
</tr>
<tr>
<td>N’-trimethyllysine</td>
<td>WG</td>
<td>C₆H₁₅N₂O₂</td>
<td>188.1525</td>
<td>84, 130, 60, 144</td>
<td>UPLC–QToF</td>
<td>[56]</td>
</tr>
<tr>
<td>glutamic acid betaine</td>
<td>WG</td>
<td>C₆H₁₀NO₄</td>
<td>189.1001</td>
<td>60, 85, 59, 58, 131, 44, 45, 103, 113</td>
<td>ESI–IT</td>
<td>[105]</td>
</tr>
<tr>
<td>histidine betaine</td>
<td>WG</td>
<td>C₆H₁₆N₂O₂</td>
<td>197.1164</td>
<td>95, 154, 60, 68, 58, 44</td>
<td>ESI–IT</td>
<td>[105]</td>
</tr>
<tr>
<td>phenylalanine betaine</td>
<td>uncertain</td>
<td>C₈H₁₄N₂O₂</td>
<td>207.1259</td>
<td>103, 107, 60, 208, 131, 79, 59, 93, 149</td>
<td>UPLC–QToF</td>
<td>[80]</td>
</tr>
<tr>
<td>tryptophan betaine</td>
<td>uncertain</td>
<td>C₈H₁₆N₂O₂</td>
<td>246.1368</td>
<td>188, 60, 146, 44, 113</td>
<td>ESI–IT</td>
<td>[105]</td>
</tr>
</tbody>
</table>

### 2.2.4 Carotenoids

Carotenoids, sometimes called tetraterpenoids, are lipophilic isoprene derivatives containing several conjugated double bonds (Figure 2.4). Chemically they can be divided into carotenes, with a pure hydrocarbon structure, and xanthophylls, which contain one or more oxygen atoms (Table 2.4). Carotenes occur widely in fruits, vegetables, and cereals, where they act as pigments, being responsible for the yellow or orange colour [118]. Although carotenoids are less concentrated in cereals compared to many other plant-based foods, the importance of cereals as a worldwide staple food makes them worth considering as a contributor to carotenoid intake [119]. Among cereals, maize contains the highest concentration of carotenoids at up to 63 µg/g. Certain wheat varieties, such as einkorn (Triticum monococcum L.), durum (Triticum turgidum subsp. durum), and emmer wheat (Triticum turgidum subsp. dicoccum), have been reported to have 2 to 4 times higher levels of carotenoids compared to common wheat [118, 119]. Oats is cited as a species with low carotenoid levels, while the information on rye is lacking [119]. Zeaxanthin is the most abundant carotenoid in maize, while lutein is the predominant form in most other cereals.

**Biological activity.** Carotenoids are relatively well studied for their physiological effects. Because they are a minor group of phytochemicals in cereals, the effects are reviewed here only briefly. One of the main functions is the provitamin A activity of e.g. β-carotene and β-cryptoxanthin [120]. In primates, lutein and zeaxanthin are concentrated in the macula lutea of the retina, and their intake from food is inversely correlated with macular degeneration and cataract [119]. Carotenoids are potent scavengers of reactive oxygen species (ROS), which is one possible mechanism to explain the reduced risk of diseases mediated by ROS, such as lung cancer in non-smokers, prostate cancer, oral and throat cancers, and cardiovascular disease [121]. There is evidence that similar protective effect occurs even when carotenoids are administered as supplements; however, the results are not fully consistent, and more
research is required in order to establish the role of these compounds in health and disease [121].

**Analysis.** Because of their high lipophilicity, limited stability, and low concentration in cereals, carotenoids are among the challenging groups of whole grain phytochemicals from the analytical perspective. In LC–MS, one difficulty is the need for certain additives and solvents unsuitable for liquid chromatography, and in GC–MS, the detection of carotenoids is limited by their poor volatility [122]. However, it is still possible to successfully extract and analyse these compounds with conventional LC–MS solvents, such as acetonitrile or 80% aqueous methanol as a solvent, and certain columns, including C30 and cyano-phase columns with low hydrophobic selectivity [118, 122, 123]. When coupled with mass spectrometry, supercritical fluid chromatography (SFC), using e.g. carbon dioxide as the mobile phase, is one suitable method for the separation of carotenoids and other lipophilic metabolites [124]. For untargeted metabolomics platforms, carotenoids are often beyond the analytical coverage of the methods, which are generally optimized for semi-polar compounds [76]. Table 2.4 presents the mass spectral characteristics of carotenoids detected in whole grains.

![Chemical structures of carotenoids detected in whole grains.](image)

**Figure 2.4.** Chemical structures of carotenoids detected in whole grains.
Table 2.4. Mass spectral characteristics of carotenoids detected in (or after the consumption of) whole grains. The origin signifies whether the compound is present in intact whole grain (WG) or as a microbial/endogenous metabolite. MM = monoisotopic neutral molecular mass in Daltons.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Origin</th>
<th>Mol. formula</th>
<th>MM</th>
<th>Fragments m/z (+)</th>
<th>Detection method</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>β-carotene</td>
<td>WG</td>
<td>C_{40}H_{56}</td>
<td>536.4382</td>
<td>444, 495, 417, 413</td>
<td>HPLC–MS</td>
<td>[123]</td>
</tr>
<tr>
<td>β-cryptoxanthin</td>
<td>WG</td>
<td>C_{40}H_{56}O</td>
<td>552.4331</td>
<td>553</td>
<td>HPLC–MS</td>
<td>[118]</td>
</tr>
<tr>
<td>lutein</td>
<td>WG</td>
<td>C_{40}H_{56}O₂</td>
<td>568.8714</td>
<td>551, 533</td>
<td>HPLC–MS</td>
<td>[118]</td>
</tr>
<tr>
<td>zeaxanthin</td>
<td>WG</td>
<td>C_{40}H_{56}O₂</td>
<td>568.8714</td>
<td>569</td>
<td>HPLC–MS</td>
<td>[118]</td>
</tr>
</tbody>
</table>

2.2.5 Flavonoids

Flavonoids are a large and diverse group of polyphenols, consisting of a C₆–C₃–C₆ skeleton (benzene ring fused into a heterocyclic ring, which contains another benzene substituent at 2' position). The six main subclasses of flavonoids are anthocyanidins, flavanols (flavan-3-ols), flavanones, flavones, flavonols, and isoflavones [125]. Flavanols can form oligomers from several units, in which case they are called proanthocyanidins or condensed tannins. The vast majority of flavonoids are stored in plants as glycosides, the most common carbohydrate moieties being arabinose, galactose, D-glucose, glucorhamnose, and L-rhamnose [126]. Unlike most plants, which produce mainly flavonoid O-glycosides, cereals predominantly synthetize C-glycosides, which are more resistant to hydrolysis [127].

Over 6500 different flavonoids have been characterized in plants [128]. They exist in cereal grains as well: more than 20 different flavonoids, excluding various glycosides, have been identified in rye alone [98]. However, the domestication of cereals involved favouring light-coloured grains, which reduced the amount of proanthocyanidins and other flavonoids, responsible for the blue, red or yellow tone, in the most commonly used cereals [129]. Pihlava et al. reported the total flavonoid content of whole-grain rye to be 46 µg/g DM and rye bran 205 µg/g DM [98], which is low compared to berries and fruits, such as common bilberry (Vaccinium myrtillus L.), with a total flavonoid content of 1900 µg/g FW [130]. Some strongly pigmented cereal varieties, such as blue wheat, pigmented maize, and black and red rice, have high concentration of anthocyanins up to 3276 µg/g FW in black rice, 1277 µg/g FW in purple corn, and 212 µg/g FW in blue wheat [131]. Therefore, due to the high amount of cereals consumed worldwide, they can be considered as an important source of flavonoids especially in regions where berries and other flavonoid-rich foods are consumed less.

Biological activity. Flavonoids are among the most studied food constituents. They were coined as “vitamin P” as early as 1930s by Nobel prize winner Albert Szent-Györgyi (the use of the term was later discontinued), and their research has increased rapidly since the 1990s [132]. Flavonoid intake has been associated with reduced mortality from cardiovascular disease and with improved markers of atherosclerosis risk, such as reduced oxidation of low-density lipoproteins (LDL), reduced
platelet aggregation, increased vasodilatation, and attenuated inflammatory response [125, 133, 134]. According to epidemiological studies, flavonoid-rich diets may reduce the risk of several types of cancer, including breast, colon, lung, pancreas, and prostate cancers [135]. However, the evidence is contradictory and limited by the difficulty of assessing the intake of the various flavonoid groups [135, 136]. Other observed or proposed activities include antiallergic, antidiabetic, antimicrobial, and gastro- and hepatoprotective effects [133]. It was first thought that flavonoids exert their biological effects via conventional antioxidant activity; however, according to the currently prevailing hypothesis, the effects are mediated via target enzymes, such as protein kinases involved in redox signalling [132]. Moreover, flavonoids can work in synergy with ascorbic acid (vitamin C) [132], caffeine [137] and possibly with each other [133] in producing their potential health effects. In conclusion, despite of the extensive research dedicated for flavonoids, it can only be established that flavonoids contribute to the beneficial health effects of whole foods and they seem to have effects of their own; mechanistic studies, dose-response studies and long-term randomized controlled trials with reliable means to measure the intake of flavonoid subclasses are warranted in the future.

**Analysis.** In general, LC–MS is well suited for the analysis of flavonoids because of the semi-polar nature of the compounds and the lack of need for extensive sample preparation techniques beyond simple homogenization, liquid extraction and filtration [138]. However, the solvent should include an acidic modifier (e.g. formic acid), which improves the retention and separation of the molecules. The most common solvents used in the extraction of flavonoids include aqueous methanol and acetonitrile [139, 140]. Anthocyanins suffer from low stability, and therefore moderate column temperature (e.g. 50 °C) will ensure that no degradation occurs before detection [138]. Flavonoid glycosides are not easily detectable in GC–MS because of their low volatility. Table 2.5 reviews the whole grain flavonoids and their structures and mass spectral characteristics.

In LC–MS, anthocyanins produce a molecular ion [M]+ due to their inherent positive charge in the heterocyclic oxygen [131]. Flavonoid glycosides exhibit a neutral loss of 162 Da (hexose unit), leaving a fragment corresponding to the m/z of the flavonoid aglycone. The glycosides may also experience in-source fragmentation even with soft ionization techniques, such as electrospray ionization (ESI); the resulting flavonoid aglycone created within the instrument can be distinguished from a natural aglycone existing in the sample by comparison of the retention time and the extracted ion chromatogram (EIC), which correspond with the peak of the glycoside in case of an in-source fragment (Figure 2.5). ESI-MS methods cannot completely distinguish between the different hexose species (glucose etc.) or whether they are O- or C-linked to the flavonoid unit [131]. However, the peaks may be separated in liquid chromatography, which makes retention time comparison with authentic standards possible (although acquiring a sufficient number of such standards may not be feasible). Geng et al. have successfully characterized various C-glycosylated flavones in wheat by combining multi-stage high resolution mass spectrometry with mass defect
filtering, providing mass spectra that are specific enough to distinguish between the compounds [141]. Due to the vast number of flavonoids with many of them having the same molecular formula, the annotation of flavonoids in untargeted analyses should always be confirmed with comparison of retention time and/or MS/MS spectra with spectral libraries or authentic standards (if available), or with purification and subsequent NMR analysis for structural elucidation.

Figure 2.5. Two overlaid extracted ion chromatograms (EICs) at m/z 433 (corresponding to apigenin 7-O-glucoside) and m/z 271 (corresponding to several flavonoid aglycones) from a phytochemical standard mixture [76] analysed with UPLC–QTOF-MS. At m/z 433, one major peak is appearing at RT 5.5 min, identified as apigenin 7-O-glucoside. The two major peaks at m/z 271 eluting at 5.5 and 6.6 min could be annotated as apigenin based on the exact mass alone; however, the peak at 5.5 min is an in-source fragment produced within the mass spectrometer from apigenin 7-O-glucoside, and the peak at 6.6 min is genistein, another flavonoid with the same exact mass as apigenin, distinguished from apigenin by its characteristic MS/MS spectrum. Apigenin was not present as a separate standard in the sample, although it could have existed there as a degradation product or impurity. Image: MS-DIAL 3.20 [142]

Table 2.5 (pages 42–44). Mass spectral characteristics of flavonoids detected in whole grains. Because flavonoids may be metabolized into phenolic acids, the metabolites are presented in Table 2.7. MM = monoisotopic neutral molecular mass in Daltons. The various flavonoid glycosides reported by Geng et al. (n = 72) [141] and Pihlava et al. (n = 53) [98] are excluded from the table due to their vast number.
Figure 2.6. Chemical structures of flavonoids detected in (or after the consumption of) whole grains. Most common flavonoid metabolites are presented.
<table>
<thead>
<tr>
<th>Compound</th>
<th>Mol. formula</th>
<th>MM</th>
<th>Fragments ( m/z (\text{-}) )</th>
<th>Fragments ( m/z (\text{+}) )</th>
<th>Detection method</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>apigenin</td>
<td>C_{15}H_{10}O_{5}</td>
<td>270.0528</td>
<td></td>
<td></td>
<td>HPLC–TOF-MS</td>
<td>[143]</td>
</tr>
<tr>
<td>naringenin</td>
<td>C_{15}H_{10}O_{5}</td>
<td>272.0685</td>
<td></td>
<td></td>
<td>UPLC–QTOF-MS</td>
<td>[98]</td>
</tr>
<tr>
<td>kaempferol</td>
<td>C_{15}H_{10}O_{6}</td>
<td>286.0477</td>
<td></td>
<td></td>
<td>UPLC–QTOF-MS</td>
<td>[98]</td>
</tr>
<tr>
<td>luteolin</td>
<td>C_{15}H_{10}O_{6}</td>
<td>286.0477</td>
<td></td>
<td></td>
<td>UPLC–QTOF-MS</td>
<td>[98]</td>
</tr>
<tr>
<td>cyanidin</td>
<td>C_{15}H_{10}O_{6}^+</td>
<td>287.0556</td>
<td></td>
<td></td>
<td>UPLC–QTOF-MS</td>
<td>[98]</td>
</tr>
<tr>
<td>(+)-catechin</td>
<td>C_{15}H_{12}O_{6}</td>
<td>290.0790, 245, 221, 203, 137</td>
<td>329, 313, 255, 237, 139</td>
<td></td>
<td>HPLC–ESI-MS</td>
<td>[144]</td>
</tr>
<tr>
<td>chrysoeriol</td>
<td>C_{15}H_{10}O_{6}</td>
<td>300.0634</td>
<td></td>
<td></td>
<td>UPLC–QTOF-MS</td>
<td>[98]</td>
</tr>
<tr>
<td>quercetin</td>
<td>C_{15}H_{10}O_{7}</td>
<td>302.0427</td>
<td></td>
<td></td>
<td>UPLC–QTOF-MS</td>
<td>[98]</td>
</tr>
<tr>
<td>delphinidin</td>
<td>C_{15}H_{10}O_{7}^+</td>
<td>303.0505</td>
<td></td>
<td></td>
<td>UPLC–QTOF-MS</td>
<td>[98]</td>
</tr>
<tr>
<td>rhamnetin</td>
<td>C_{15}H_{10}O_{7}</td>
<td>316.0583</td>
<td></td>
<td></td>
<td>UPLC–QTOF-MS</td>
<td>[98]</td>
</tr>
<tr>
<td>petunidin</td>
<td>C_{15}H_{12}O_{7}^+</td>
<td>317.0661</td>
<td></td>
<td></td>
<td>UPLC–QTOF-MS</td>
<td>[98]</td>
</tr>
<tr>
<td>tricin</td>
<td>C_{17}H_{14}O_{7}</td>
<td>330.0740</td>
<td></td>
<td></td>
<td>HPLC–TOF-MS</td>
<td>[143]</td>
</tr>
<tr>
<td>laricitrin</td>
<td>C_{15}H_{12}O_{6}</td>
<td>332.0532</td>
<td></td>
<td></td>
<td>UPLC–QTOF-MS</td>
<td>[98]</td>
</tr>
<tr>
<td>dihydrotricin</td>
<td>C_{17}H_{14}O_{7}</td>
<td>332.0896</td>
<td></td>
<td></td>
<td>UPLC–QTOF-MS</td>
<td>[98]</td>
</tr>
<tr>
<td>6-prenylnaringenin</td>
<td>C_{20}H_{20}O_{5}</td>
<td>340.1311, 289, 183, 307, 321</td>
<td></td>
<td></td>
<td>LC–LTQ-Orbitrap</td>
<td>[145]</td>
</tr>
<tr>
<td>syringetin</td>
<td>C_{17}H_{14}O_{6}</td>
<td>346.0689</td>
<td></td>
<td></td>
<td>UPLC–QTOF-MS</td>
<td>[98]</td>
</tr>
<tr>
<td>licoisoflavone A</td>
<td>C_{20}H_{20}O_{6}</td>
<td>354.1103, 235, 320, 255</td>
<td></td>
<td></td>
<td>LC–LTQ-Orbitrap</td>
<td>[145]</td>
</tr>
<tr>
<td>5-hydroxy-3,6,7,4′-tetrámethoxyflavone</td>
<td>C_{15}H_{10}O_{7}</td>
<td>358.1053</td>
<td></td>
<td></td>
<td>LC–LTQ-Orbitrap</td>
<td>[145]</td>
</tr>
<tr>
<td>pelargonidin 3-arabinoside</td>
<td>C_{20}H_{12}O_{6}^+</td>
<td>403.1029</td>
<td>271</td>
<td></td>
<td>UPLC–API-MS</td>
<td>[146]</td>
</tr>
<tr>
<td>cyanidin 3-arabinoside</td>
<td>C_{20}H_{12}O_{6}^+</td>
<td>419.0978</td>
<td>287</td>
<td></td>
<td>UPLC–API-MS</td>
<td>[146]</td>
</tr>
<tr>
<td>vitexin</td>
<td>C_{21}H_{20}O_{10}</td>
<td>432.1056, 311, 341</td>
<td></td>
<td></td>
<td>UPLC–LTQ-Orbitrap</td>
<td>[129]</td>
</tr>
<tr>
<td>isovitexin</td>
<td>C_{21}H_{20}O_{10}</td>
<td>432.1056, 311, 341, 413</td>
<td></td>
<td></td>
<td>UPLC–LTQ-Orbitrap</td>
<td>[129]</td>
</tr>
<tr>
<td>kaempferol 3-O-rhamnoside</td>
<td>C_{21}H_{20}O_{10}</td>
<td>432.1056, 153, 171, 399, 385</td>
<td></td>
<td></td>
<td>LC–LTQ-Orbitrap</td>
<td>[145]</td>
</tr>
<tr>
<td>pelargonidin 3-glucoside</td>
<td>C_{21}H_{20}O_{10}^+</td>
<td>433.1135</td>
<td>271</td>
<td></td>
<td>UPLC–API-MS</td>
<td>[146]</td>
</tr>
<tr>
<td>pelargonidin 3-galactoside</td>
<td>C_{21}H_{20}O_{10}^+</td>
<td>433.1135</td>
<td>287</td>
<td></td>
<td>UPLC–API-MS</td>
<td>[146]</td>
</tr>
<tr>
<td>naringenin 7-glucoside</td>
<td>C_{21}H_{20}O_{10}</td>
<td>434.1212, 313, 253, 310, 231</td>
<td></td>
<td></td>
<td>LC–LTQ-Orbitrap</td>
<td>[145]</td>
</tr>
<tr>
<td>delphinidin 3-arabinoside</td>
<td>C_{20}H_{10}O_{11}^+</td>
<td>435.0927</td>
<td>303</td>
<td></td>
<td>UPLC–API-MS</td>
<td>[146]</td>
</tr>
<tr>
<td>formononetin (glycosylated and methylated)</td>
<td>C_{23}H_{20}O_{9}</td>
<td>444.1420</td>
<td></td>
<td></td>
<td>HPLC–TOF-MS</td>
<td>[143]</td>
</tr>
<tr>
<td>orientin/isorientin</td>
<td>C_{21}H_{20}O_{11}</td>
<td>448.1006</td>
<td></td>
<td></td>
<td>HPLC–TOF-MS</td>
<td>[143]</td>
</tr>
<tr>
<td>luteolin 7-O-hexoside (cyanoside)</td>
<td>C_{21}H_{20}O_{11}</td>
<td>448.1006, 285</td>
<td></td>
<td></td>
<td>HPLC–QTOF-MS</td>
<td>[147]</td>
</tr>
<tr>
<td>cyanidin 3-galactoside</td>
<td>C_{21}H_{20}O_{11}^+</td>
<td>449.1084</td>
<td>287</td>
<td></td>
<td>UPLC–API-MS</td>
<td>[146]</td>
</tr>
</tbody>
</table>

42
<table>
<thead>
<tr>
<th>Compound</th>
<th>Mol. formula</th>
<th>MM</th>
<th>Fragments m/z (−)</th>
<th>Fragments m/z (+)</th>
<th>Detection method</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>cyanidin 3-glucoside</td>
<td>C_{21}H_{27}O_{11}</td>
<td>449.1084</td>
<td></td>
<td>287</td>
<td>UPLC-API-MS</td>
<td>[146]</td>
</tr>
<tr>
<td>chrysoeriol 6-hexoside</td>
<td>C_{23}H_{25}O_{11}</td>
<td>462.1162</td>
<td></td>
<td>445, 397, 427, 367, 343, 409, 313</td>
<td>HPLC-API-IT-MS</td>
<td>[148]</td>
</tr>
<tr>
<td>peonidin 3-glucoside</td>
<td>C_{22}H_{23}O_{11}</td>
<td>463.1240</td>
<td></td>
<td>301</td>
<td>UPLC-API-MS</td>
<td>[146]</td>
</tr>
<tr>
<td>delphinidin 3-galactoside</td>
<td>C_{21}H_{25}O_{12}</td>
<td>465.1033</td>
<td></td>
<td>303</td>
<td>UPLC-API-MS</td>
<td>[146]</td>
</tr>
<tr>
<td>petunidin 3-glucoside</td>
<td>C_{23}H_{25}O_{12}</td>
<td>479.1190</td>
<td></td>
<td>317</td>
<td>UPLC-API-MS</td>
<td>[146]</td>
</tr>
<tr>
<td>3,7-dimethylquercetin glycosylate</td>
<td>C_{25}H_{26}O_{12}</td>
<td>492.1268</td>
<td></td>
<td></td>
<td>HPLC-TOF-MS</td>
<td>[143]</td>
</tr>
<tr>
<td>tricin O-hexoside</td>
<td>C_{23}H_{26}O_{12}</td>
<td>492.1268</td>
<td></td>
<td>331</td>
<td>HPLC-API-IT-MS</td>
<td>[148]</td>
</tr>
<tr>
<td>malvidin 3-glucoside</td>
<td>C_{23}H_{26}O_{12}</td>
<td>493.1346</td>
<td></td>
<td>331</td>
<td>UPLC-API-MS</td>
<td>[146]</td>
</tr>
<tr>
<td>3,7-dimethylquercetin (glycosylated and acetylated)</td>
<td>C_{23}H_{26}O_{13}</td>
<td>534.1373</td>
<td>353, 383, 443, 413, 473, 515</td>
<td></td>
<td>HPLC-TOF-MS</td>
<td>[143]</td>
</tr>
<tr>
<td>apigenin-6,8-di-C-pentoside</td>
<td>C_{23}H_{26}O_{13}</td>
<td>534.1373</td>
<td>353, 383, 443, 413, 473, 515</td>
<td></td>
<td>HPLC-QTOF-MS</td>
<td>[147]</td>
</tr>
<tr>
<td>hemsleyanoside</td>
<td>C_{23}H_{26}O_{13}</td>
<td>548.1529</td>
<td></td>
<td></td>
<td>LC-LTQ-Orbitrap</td>
<td>[145]</td>
</tr>
<tr>
<td>apigenin-6-C-arabinoside-8-C-hexoside (schaftoside/isoschaftoside)</td>
<td>C_{25}H_{26}O_{14}</td>
<td>564.1479</td>
<td></td>
<td></td>
<td>HPLC-TOF-MS</td>
<td>[143]</td>
</tr>
<tr>
<td>apigenin-6-C-pentoside-8-C-hexoside</td>
<td>C_{25}H_{26}O_{14}</td>
<td>564.1479</td>
<td>353, 383, 443, 413, 473, 503, 545</td>
<td></td>
<td>HPLC-QTOF-MS</td>
<td>[147]</td>
</tr>
<tr>
<td>procyanidin B2</td>
<td>C_{25}H_{26}O_{12}</td>
<td>578.1424</td>
<td>425, 407, 289, 617, 601, 579, 427, 291</td>
<td></td>
<td>HPLC-ESI-MS</td>
<td>[144]</td>
</tr>
<tr>
<td>procyanidin B3</td>
<td>C_{25}H_{26}O_{12}</td>
<td>578.1424</td>
<td></td>
<td></td>
<td>HPLC-TOF-MS</td>
<td>[143]</td>
</tr>
<tr>
<td>isovitexin 2&quot;-O-rhamnose</td>
<td>C_{27}H_{30}O_{14}</td>
<td>578.1636</td>
<td></td>
<td></td>
<td>HPLC-TOF-MS</td>
<td>[143]</td>
</tr>
<tr>
<td>apigenin 7-O-neohesperidoside</td>
<td>C_{27}H_{30}O_{14}</td>
<td>578.1636</td>
<td>269</td>
<td></td>
<td>HPLC-QTOF-MS</td>
<td>[147]</td>
</tr>
<tr>
<td>lucenin-1/3 (luteolin C-xylloside-C-glucoside)</td>
<td>C_{26}H_{26}O_{15}</td>
<td>580.1428</td>
<td></td>
<td></td>
<td>HPLC-TOF-MS</td>
<td>[143]</td>
</tr>
<tr>
<td>auriculatin 4&quot;-O-glucoside</td>
<td>C_{27}H_{30}O_{11}</td>
<td>582.2101</td>
<td></td>
<td></td>
<td>LC-LTQ-Orbitrap</td>
<td>[145]</td>
</tr>
<tr>
<td>prodelphinidin B3</td>
<td>C_{25}H_{26}O_{13}</td>
<td>594.1373</td>
<td>467, 303, 289, 617, 443, 427, 289</td>
<td></td>
<td>HPLC-ESI-MS</td>
<td>[144]</td>
</tr>
<tr>
<td>prodelphinidin dimer (GC–C)</td>
<td>C_{25}H_{26}O_{13}</td>
<td>594.1373</td>
<td>467, 289, 633, 617, 443, 427, 291</td>
<td></td>
<td>HPLC-ESI-MS</td>
<td>[144]</td>
</tr>
<tr>
<td>kaempferol 3-rhamnoside-7-glucoside</td>
<td>C_{21}H_{23}O_{15}</td>
<td>594.1584</td>
<td></td>
<td></td>
<td>LC-LTQ-Orbitrap</td>
<td>[145]</td>
</tr>
<tr>
<td>apigenin 6,8-di-C-hexoside (vicein 2)</td>
<td>C_{27}H_{30}O_{15}</td>
<td>594.1585</td>
<td>353, 383, 443, 503, 575</td>
<td></td>
<td>HPLC-QTOF-MS</td>
<td>[147]</td>
</tr>
<tr>
<td>chrysoeriol 6-C-pentoside-8-C-hexoside</td>
<td>C_{27}H_{30}O_{15}</td>
<td>594.1585</td>
<td>383, 413, 473, 503, 575, 443, 533</td>
<td></td>
<td>HPLC-QTOF-MS</td>
<td>[147]</td>
</tr>
<tr>
<td>kaempferol 7-O-neohesperidoside</td>
<td>C_{27}H_{30}O_{15}</td>
<td>594.1585</td>
<td>285</td>
<td></td>
<td>HPLC-QTOF-MS</td>
<td>[147]</td>
</tr>
<tr>
<td>luteolin 7-O-neohesperidoside</td>
<td>C_{27}H_{30}O_{15}</td>
<td>594.1585</td>
<td>285</td>
<td></td>
<td>HPLC-QTOF-MS</td>
<td>[147]</td>
</tr>
<tr>
<td>methylisoorientin 2&quot;-O-rhamnoside</td>
<td>C_{29}H_{32}O_{15}</td>
<td>608.1741</td>
<td></td>
<td></td>
<td>HPLC-TOF-MS</td>
<td>[143]</td>
</tr>
</tbody>
</table>

43
Lignans are polyphenols formed from a dibenzylbutane skeleton, which usually contains two C8–C8′-linked phenylpropanoid units (Figure 2.7). Based on the more specific structure, they can be further classified into aryltetralines, butanediols, dibenzylbutyrolactones, furanolignans, and furofuranolignans [150]. In case of one or two additional phenylpropanoid units, these lignans are referred to as sesquilignans and dilignans, respectively [151]. Lignans are widespread within Plantae [150] and they
exist in all the major cereals [152, 153]. Rye is a rich source of lignans, having a concentration of 19 µg/g in the whole grain and 50–70 µg/g in the bran section [152]. The most abundant lignan species in rye, in descending order of their concentration, are syringaresinol, pinoresinol, lariciresinol, hydroxymatairesinol, and medioresinol [152]. Wheat is another major source of these compounds, while triticale, oats, spelt, rice, and barley contain them in smaller quantities. As with many other bioactives compounds, lignans are concentrated in the outer layers of the cereal grain [152]. Although flaxseeds and sesame seeds are by far the richest sources of lignans by concentration [154], their consumption in the Western diet is relatively low, emphasizing the importance of whole grains as a lignan source.

**Biological activity.** Perhaps owing to their wide distribution in plant-based foods, the health effects of lignans have been studied relatively widely compared to many other classes of phytochemicals. Lignans are plant oestrogens, their structures resembling that of 17β-oestradiol, and they may act as oestrogen agonists or antagonists; these effects have been observed both in vitro and in vivo [155]. This has been shown also for enterolignans, the gut microbial metabolites of lignans [156, 157]; in fact, the microbial transformation of lignans even seems to increase their oestrogenic activity. Acting as natural ligands for the oestrogen receptor (ER), they can exert anticarcinogenic effects by either competing with oestradiol in binding to ER or by activating transcriptional coregulators by ER activation [155]. ERs are abundant in the colon, and their reduction is related to cancer pathogenesis [158]. Enterolignans have been shown to protect against colon cancer in epidemiological and case–control studies [159, 160]; the likely mechanism is the inhibition of tumour cell growth by cytostatic and apoptotic processes [161]. There is some evidence of a protective effect against prostate cancer as well [155]. Although lignans have been associated with a decreased risk of breast cancer, the findings are inconsistent and their consumption as supplements should be recommended with caution due to oestrogen-dependent stimulation of certain tumours.

Another potential health impact of lignans is based on their ability to act as antioxidants. Prasad (2000) reported that secoisolariciresinol diglucoside (SDG) had antioxidant potency 1.3 times greater than vitamin E, while its metabolites secoisolariciresinol, enterodiol, and enterolactone had a potency 4.9, 5.0, and 4.4 times that of vitamin E, respectively [162]. Flaxseed may protect against cardiovascular disease by prevention atherosclerotic plaque formation, and SDG is hypothesized as the main component within flaxseed behind the effect [163]. Lignans can also inhibit cholesterol absorption and provide additional benefits in the prevention of CVD. Evidence of the role of lignans in CVD prevention was also obtained from an association between high urinary concentration of enterolactone and lower all-cause and cardiovascular mortality [164]. Finally, lignans have shown renal and hepatoprotective effects in animal experiments [155], but as a limitation, they were administered as part of flaxseed, which introduces potential confounding factors, such as polyunsaturated fatty acids and dietary fiber. In case of cereals, where lignans are accompanied by
several other major groups of phytochemicals, associating the observed effects with lignans themselves would be even more challenging.

**Analysis.** Traditionally, lignans have been analysed with GC or GC–MS with derivatization; however, advances LC–MS have increased its popularity particularly for the qualitative characterization of lignans in plant or mammalian body fluid samples [162]. In LC–MS, lignans can be detected in both positive and negative ionization mode, although negative mode has been used more often in the identification (Table 2.6). Lignans do not lend themselves easily for detection in plant samples due to them being bound in the cell wall matrix. The extraction process usually requires several steps, including subsequent extraction with a non-polar solvent (e.g. hexane) and a more polar solvent (e.g. methanol) and in some cases acid, alkaline or enzymatic hydrolysis [151]. This produces a challenge for untargeted metabolomics, where such targeted extraction methods are not always feasible.

Different lignan classes produce distinctive fragmentation patterns in the MS/MS spectra. Butanediols exhibit a neutral loss of 48 Da from the simultaneous loss of H₂O and formaldehyde [150]. With dibenzylbutyrolactones, a neutral loss of 15 Da from the cleavage of a methoxy radical and 44 Da from CO₂ may be observed [150, 165]. Furofuranolignans cleave from within the tetrahydrofuran ring, producing negatively charged fragment ions at \( m/z \) 151 (for pinoresinol) and \( m/z \) 181 (for syringaresinol), which may further lose a methoxy radical, resulting in fragments at \( m/z \) 136 and 166, respectively [165]. Sesquilignans and dilignans produce a fragment ion at \( m/z \) 195, originating from the additional phenylpropanoid unit [165]. Lignans with a hydroxymethyl group, such as butanediols and tetrahydrofuranolignans, typically lose a formaldehyde moiety during the collision-induced dissociation (CID), producing a neutral loss of 30 Da [165].

Figure 2.7 (next page). Chemical structures of lignans and their metabolites detected in (or after the consumption of) whole grains.

Table 2.6 (next page). Mass spectral characteristics of lignans and their metabolites. The origin signifies whether the compound is present in intact whole grain (WG) or as a microbial/endogenous metabolite. MM = monoisotopic neutral molecular mass in Daltons. * Putative structures.
<table>
<thead>
<tr>
<th>Compound</th>
<th>Origin</th>
<th>Mol. formula</th>
<th>MM</th>
<th>Fragments m/z (−)</th>
<th>Fragments m/z (+)</th>
<th>Detection method</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>enterolactone</td>
<td>microbiota</td>
<td>C_{18}H_{24}O_{8}</td>
<td>298.1205</td>
<td>253, 165, 145, 116</td>
<td></td>
<td>UPLC–QTOF</td>
<td>[79]</td>
</tr>
<tr>
<td>enterodiol</td>
<td>microbiota</td>
<td>C_{18}H_{24}O_{6}</td>
<td>302.1518</td>
<td></td>
<td></td>
<td>UPLC–QTOF</td>
<td>[79]</td>
</tr>
<tr>
<td>hinokinin</td>
<td>WG</td>
<td>C_{20}H_{18}O_{6}</td>
<td>354.1103</td>
<td></td>
<td></td>
<td>HPLC–TOF</td>
<td>[143]</td>
</tr>
<tr>
<td>pinoresinol</td>
<td>WG</td>
<td>C_{20}H_{22}O_{6}</td>
<td>358.1416</td>
<td>342, 311, 175, 151, 136</td>
<td></td>
<td>UPLC–QTOF</td>
<td>[165]</td>
</tr>
<tr>
<td>(−)-dihydrocubebin</td>
<td>WG</td>
<td>C_{20}H_{22}O_{6}</td>
<td>358.1416</td>
<td></td>
<td></td>
<td>LC–LTQ-Orbitalrap</td>
<td>[145]</td>
</tr>
<tr>
<td>matairesinol</td>
<td>WG</td>
<td>C_{20}H_{22}O_{7}</td>
<td>359.1416</td>
<td>342, 313, 298, 221, 161, 147, 137, 122</td>
<td></td>
<td>UPLC–QTOF</td>
<td>[165]</td>
</tr>
<tr>
<td>laticiresinol</td>
<td>WG</td>
<td>C_{20}H_{22}O_{6}</td>
<td>360.1573</td>
<td>329, 192, 178, 175, 160, 116</td>
<td></td>
<td>UPLC–QTOF</td>
<td>[165]</td>
</tr>
<tr>
<td>hydroxymatairesinol</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hydroxymedioreisol</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hedyotisols</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pinoresinol</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>secoisolaricresinol</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dihydrocubebin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>enterolactone</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7-hydroxyenterolactone</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>enterodiol</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## 2.2.7 Phenolic acids

Phenolic acids are among the most abundant phytochemicals in whole grains. Structurally, these aromatic carboxylic acids can be classified into two groups: hydroxybenzoic acids with a C₆–C₁ skeleton (corresponding to benzoic acid) and hydroxycinnamic acids consisting of a C₆–C₃ skeleton (cinnamic acid) (Table 2.7). p-Coumaric acid, the simplest hydroxycinnamic acid, is synthetized in plants either from tyrosine or from phenylalanine via cinnamic acid and further synthetized into caffeic, ferulic and sinapic acids [167]. Ferulic acid, the most abundant phenolic acid in cereals, exists also as a dimer in various forms depending on the location of the covalent bond between the units [168]. In addition, phenolic acids are commonly linked to other small molecules, such as polyamines (agmatine, putrescine and spermidine) [98] and phytosterols (γ-oryzanol in rice bran) [169]. Phenolic acids have been shown to be important structural elements in the cell wall of both soluble and insoluble dietary fibre, cross-linking polysaccharide constituents of plant cell walls. Diferulates are mostly responsible for the linkage of arabinoxylans, while coumarates bind lignin structures together [170]). Phenolic acids may have additional functions.
in plants, acting as defence molecules against e.g. *Fusarium* fungi, which produce several mycotoxins potentially harmful to humans as well [6]. The total content of phenolic acids, predominantly in the bound form, is relatively high in whole-grain rye (1.36 mg/g fresh weight) and wheat (1.34 mg/g FW) and lower in oats (0.47 mg/g FW) and barley (0.45 mg/g FW) [171]. According to Mattila et al., whole-grain wheat flour is 8 to 9 times more abundant in phenolic acids compared to white wheat flour.

**Biological activity.** Phenolic acids are mostly known for their antioxidant and radical scavenging activity. The antioxidant effect is based on the presence of aromatic hydroxyl groups and influenced by their position and the type of carboxylate group, as reviewed in detail by Rice-Evans et al. [172]. In a study by Sroka et al., gallic acid had the strongest scavenging activity against hydrogen peroxide (H$_2$O$_2$) and 1,1-diphenyl-2-picrylhydrazyl (DPPH$^+$) radicals, followed by caffeic and protocatechuic acid [173]. Dietary polyphenols, such as (+)-catechin and (−)-epicatechin, are precursors of phenolic acids, which makes the association of health benefits to strictly phenolic acids difficult, and many *in vivo* studies have thus focused on the overall effects of phenolics [174].

**Analysis.** Since phenolic acids are mostly bound to the fibre matrix in cereals, any quantitative analyses on the total phenolic acid content require a hydrolysis step by enzymatic processing or acid/alkaline treatment [171]. However, the content of free phenolic acids in cereal bran is still high enough for them to be detected and identified in untargeted metabolomics, where a simple extraction procedure is used [74]. Most common solvents in the extraction of phenolic acids (and other phenolics) include acetone, diethyl ether, ethanol, ethyl acetate, and methanol; however, the most polar phenolic acids require an aqueous solution of acetone, ethanol or methanol [139]. The carboxylic acid group present in phenolic acids is prone to deprotonation thus yielding a negative ion, but the compounds are detectable for the most part also in the positive ionization mode, possibly due to acidic regulators often added to the HPLC mobile phase to promote ionization [76]. In LC–MS negative mode, ferulic acid typically produces fragments at *m/z* 178, 149, and 134, resulting from demethylation, decarboxylation, and the combination of both, respectively [175]. Decarboxylation can be observed in other phenolic acids as well as a neutral loss of 44 Da, resulting in characteristic fragment ions at *m/z* 135 in caffeic acid, *m/z* 119 in *p*-coumaric acid, and *m/z* 341 in diferulic acids (Figure 2.8, Table 2.7).

Figure 2.8. (next page) Characteristic MS/MS spectra of typical phenolic acids in whole grains (except dihydrocaffeic acid, a microbial metabolite) in the negative ionization mode (−20 V) and positive ionization mode (+20 V). The spectra were acquired from reference standards utilizing UPLC–QTOF-MS. Images: MS-DIAL 3.30 [142].
**p-Coumaric acid**

**Vanillic acid**

**Caffeic acid**

**Dihydrocaffeic acid**

(not detected in ESI+)

**Ferulic acid**

**Sinapic acid**
Figure 2.9. Chemical structures of phenolic acids and their metabolites detected in (or after the consumption of) whole grains.

Table 2.7. Mass spectral characteristics of phenolic acids and their metabolites. The origin signifies whether the compound is present in intact whole grain (WG) or as a microbial/endogenous metabolite. Most of the numerous phenolic acid derivatives, including polyamines, are not included in the table due to space restrictions. MM = monoisotopic neutral molecular mass in Daltons; * may originate from other polyphenols as well; † unpublished data.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Origin</th>
<th>Mol. formula</th>
<th>MM</th>
<th>Fragments m/z (−)</th>
<th>Fragments m/z (+)</th>
<th>Detection method</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>benzoic acid*</td>
<td>microbiota</td>
<td>C₆H₅O₂</td>
<td>122.0368</td>
<td></td>
<td></td>
<td>GC×GCTOF</td>
<td>[176]</td>
</tr>
<tr>
<td>p-salicylic acid*</td>
<td>unknown</td>
<td>C₆H₅O₃</td>
<td>137.0239</td>
<td>93</td>
<td></td>
<td>UPLC–QTOF</td>
<td>[177]</td>
</tr>
<tr>
<td>p-hydroxybenzoic acid*</td>
<td>microbiota</td>
<td>C₆H₅O₃</td>
<td>138.0317</td>
<td>93, 65</td>
<td></td>
<td>UPLC–QTOF</td>
<td>[79]</td>
</tr>
<tr>
<td>m-hydroxybenzoic acid*</td>
<td>microbiota</td>
<td>C₆H₅O₃</td>
<td>138.0317</td>
<td></td>
<td></td>
<td>GC×GCTOF</td>
<td>[176]</td>
</tr>
<tr>
<td>phenylpropionic acid</td>
<td>microbiota</td>
<td>C₆H₁₀O₂</td>
<td>150.0681</td>
<td>147, 117, 105</td>
<td></td>
<td>UPLC–QTOF</td>
<td>[79]</td>
</tr>
<tr>
<td>4-hydroxyphenylacetic acid*</td>
<td>microbiota</td>
<td>C₆H₆O₂</td>
<td>152.0473</td>
<td>107, 79</td>
<td></td>
<td>UPLC–QTOF</td>
<td>[177]</td>
</tr>
<tr>
<td>protocatechuic acid*</td>
<td>microbiota</td>
<td>C₆H₅O₄</td>
<td>154.0266</td>
<td>109, 53, 108, 81</td>
<td></td>
<td>UPLC–QTOF</td>
<td>[79, 177]</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Compound</th>
<th>Origin</th>
<th>Mol. formula</th>
<th>MM</th>
<th>Fragments m/z (-)</th>
<th>Fragments m/z (+)</th>
<th>Detection method</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>m-coumaric acid</td>
<td>microbiota</td>
<td>C₆H₅O₃</td>
<td>164.0473</td>
<td>119, 163</td>
<td>119, 91, 147</td>
<td>HPLC–ESI–IT</td>
<td>[178]</td>
</tr>
<tr>
<td>p-coumaric acid</td>
<td>WG</td>
<td>C₆H₅O₃</td>
<td>164.0473</td>
<td>121, 106, 119, 103</td>
<td></td>
<td>UPLC–QTOF</td>
<td>†</td>
</tr>
<tr>
<td>3-(3-hydroxyphenyl)propionic acid*</td>
<td>microbiota</td>
<td>C₆H₇O₃</td>
<td>166.0630</td>
<td>106, 119, 121</td>
<td></td>
<td>GC×GCQTOF</td>
<td>[176]</td>
</tr>
<tr>
<td>3-(4-hydroxyphenyl)propionic acid*</td>
<td>microbiota</td>
<td>C₆H₇O₃</td>
<td>166.0630</td>
<td></td>
<td></td>
<td>UPLC–QTOF</td>
<td>[177]</td>
</tr>
<tr>
<td>3,4-dihydroxy-phenylacetic acid*</td>
<td>microbiota</td>
<td>C₆H₈O₄</td>
<td>168.0423</td>
<td></td>
<td></td>
<td>GC×GCQTOF</td>
<td>[176]</td>
</tr>
<tr>
<td>vanillic acid</td>
<td>WG</td>
<td>C₆H₆O₄</td>
<td>168.0423</td>
<td>108, 152, 91, 80</td>
<td>65, 93</td>
<td>UPLC–QTOF</td>
<td>†</td>
</tr>
<tr>
<td>hippuric acid</td>
<td>liver</td>
<td>C₈H₈NO₃</td>
<td>179.0582</td>
<td>134, 178, 77, 56</td>
<td>105, 77</td>
<td>UPLC–QTOF</td>
<td>[80]</td>
</tr>
<tr>
<td>caffeic acid</td>
<td>WG</td>
<td>C₆H₄O₄</td>
<td>180.0423</td>
<td>135, 179</td>
<td>135, 117, 163, 89*</td>
<td>UPLC–QTOF</td>
<td>†</td>
</tr>
<tr>
<td>dihydrocaffeic acid</td>
<td>sourdough,</td>
<td>C₈H₁₀O₄</td>
<td>182.0579</td>
<td>137, 135, 43, 93</td>
<td></td>
<td>UPLC–QTOF</td>
<td>[56, 177]</td>
</tr>
<tr>
<td>2-methylhippuric acid*</td>
<td>microbiota</td>
<td>C₁₀H₁₄NO₃</td>
<td>193.0739</td>
<td></td>
<td></td>
<td>HPLC–QTOF</td>
<td>[166]</td>
</tr>
<tr>
<td>ferulic acid</td>
<td>WG</td>
<td>C₁₀H₁₀O₄</td>
<td>194.0579</td>
<td>133, 134, 178</td>
<td>117, 145, 149, 89*</td>
<td>UPLC–QTOF</td>
<td>[7]</td>
</tr>
<tr>
<td>dihydroferulic acid</td>
<td>sourdough,</td>
<td>C₁₀H₁₂O₄</td>
<td>196.0736</td>
<td>136, 121, 45, 135</td>
<td></td>
<td>UPLC–QTOF</td>
<td>[56, 79]</td>
</tr>
<tr>
<td>syringic acid</td>
<td>WG</td>
<td>C₆H₁₀O₅</td>
<td>198.0528</td>
<td></td>
<td></td>
<td>HPLC–TOF</td>
<td>[143]</td>
</tr>
<tr>
<td>benzoic acid sulfate*</td>
<td>liver</td>
<td>C₆H₅O₃</td>
<td>201.9936</td>
<td>137, 93</td>
<td></td>
<td>HPLC–ESI–IT</td>
<td>[178]</td>
</tr>
<tr>
<td>cinnamoylglycine</td>
<td>unknown</td>
<td>C₁₁H₁₂NO₅</td>
<td>205.0739</td>
<td>160, 130, 204, 103</td>
<td>131, 103, 77</td>
<td>UPLC–QTOF</td>
<td>[80]</td>
</tr>
<tr>
<td>sinapic acid</td>
<td>WG</td>
<td>C₁₁H₁₂O₅</td>
<td>224.0685</td>
<td>193, 93, 121, 149</td>
<td>147, 175, 119, 94*</td>
<td>UPLC–QTOF</td>
<td>[7]</td>
</tr>
<tr>
<td>dihydroxysinapic acid</td>
<td>sourdough,</td>
<td>C₁₁H₁₄O₅</td>
<td>226.0841</td>
<td>166, 151</td>
<td></td>
<td>UPLC–QTOF</td>
<td>[56, 79]</td>
</tr>
<tr>
<td>3-phenylpropionic acid sulfate</td>
<td>liver</td>
<td>C₆H₅O₃</td>
<td>230.0249</td>
<td></td>
<td></td>
<td>HPLC–ESI–IT</td>
<td>[178]</td>
</tr>
<tr>
<td>protocatechuic acid sulfate*</td>
<td>liver</td>
<td>C₇H₆O₇</td>
<td>233.9834</td>
<td>153, 233, 151, 97</td>
<td></td>
<td>UPLC–QTOF</td>
<td>[80]</td>
</tr>
<tr>
<td>coumaric acid sulfate*</td>
<td>liver</td>
<td>C₆H₅O₃</td>
<td>244.0042</td>
<td></td>
<td></td>
<td>HPLC–ESI–IT</td>
<td>[178]</td>
</tr>
<tr>
<td>vanillic acid sulfate</td>
<td>liver</td>
<td>C₆H₅O₃</td>
<td>247.9991</td>
<td></td>
<td></td>
<td>HPLC–ESI–IT</td>
<td>[178]</td>
</tr>
<tr>
<td>caffeic acid sulfate</td>
<td>liver</td>
<td>C₆H₆O₃</td>
<td>259.9991</td>
<td>179, 135, 97, 135</td>
<td></td>
<td>UPLC–QTOF</td>
<td>[80]</td>
</tr>
<tr>
<td>ferulic acid sulfate</td>
<td>liver</td>
<td>C₁₀H₁₀O₅</td>
<td>274.0147</td>
<td>193, 178, 149, 134</td>
<td></td>
<td>UPLC–QTOF</td>
<td>[80]</td>
</tr>
<tr>
<td>dihydroferulic acid sulfate</td>
<td>liver</td>
<td>C₁₀H₁₂O₅</td>
<td>276.0304</td>
<td>177, 123, 195, 133</td>
<td></td>
<td>UPLC–QTOF</td>
<td>[80]</td>
</tr>
<tr>
<td>p-coumaroyl pentose</td>
<td>WG</td>
<td>C₄H₈O₇</td>
<td>296.0896</td>
<td>235, 163</td>
<td>335, 319</td>
<td>HPLC–ESI–MS</td>
<td>[179]</td>
</tr>
<tr>
<td>sinapic acid sulfate</td>
<td>liver</td>
<td>C₁₁H₁₂O₅</td>
<td>304.0253</td>
<td>223, 208, 179, 164</td>
<td></td>
<td>HPLC–ESI–IT</td>
<td>[178]</td>
</tr>
</tbody>
</table>

52
<table>
<thead>
<tr>
<th>Compound</th>
<th>Origin</th>
<th>Mol. formula</th>
<th>MM</th>
<th>Fragments ( m/z ) ((-))</th>
<th>Fragments ( m/z ) ((+))</th>
<th>Detection method</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>caffeoyl pentose</td>
<td>WG</td>
<td>C_{14}H_{16}O_{5}</td>
<td>312.0845</td>
<td>179</td>
<td>351, 335, 181</td>
<td>HPLC–ESI–MS</td>
<td>[179]</td>
</tr>
<tr>
<td>feruloyl pentose</td>
<td>WG</td>
<td>C_{15}H_{18}O_{5}</td>
<td>326.1002</td>
<td>193, 175</td>
<td>365, 349</td>
<td>HPLC–ESI–MS</td>
<td>[179]</td>
</tr>
<tr>
<td>(p)-coumaroyl hexose</td>
<td>WG</td>
<td>C_{15}H_{18}O_{5}</td>
<td>326.1002</td>
<td>163, 145</td>
<td>365</td>
<td>HPLC–ESI–MS</td>
<td>[179]</td>
</tr>
<tr>
<td>3-hydroxyphenylpropionic acid glucuronide</td>
<td>liver</td>
<td>C_{15}H_{18}O_{5}</td>
<td>342.0951</td>
<td>181, 137</td>
<td></td>
<td>HPLC–ESI–IT</td>
<td>[178]</td>
</tr>
<tr>
<td>caffeoyl hexose</td>
<td>WG</td>
<td>C_{15}H_{18}O_{5}</td>
<td>342.0951</td>
<td>179, 161</td>
<td>365</td>
<td>HPLC–ESI–MS</td>
<td>[179]</td>
</tr>
<tr>
<td>homovanillic acid glucuronide*</td>
<td>liver</td>
<td>C_{15}H_{18}O_{10}</td>
<td>358.0900</td>
<td></td>
<td></td>
<td>HPLC–QTOF</td>
<td>[166]</td>
</tr>
<tr>
<td>ferulic acid glucuronide</td>
<td>liver</td>
<td>C_{15}H_{18}O_{10}</td>
<td>370.0900</td>
<td></td>
<td></td>
<td>HPLC–QTOF</td>
<td>[166]</td>
</tr>
<tr>
<td>5,5'-diferulic acid</td>
<td>WG</td>
<td>C_{20}H_{18}O_{5}</td>
<td>386.1002</td>
<td>341, 297</td>
<td></td>
<td>HPLC–ESI–MS</td>
<td>[180]</td>
</tr>
<tr>
<td>8,5'-diferulic acid</td>
<td>WG</td>
<td>C_{20}H_{18}O_{5}</td>
<td>386.1002</td>
<td>341, 326, 370, 283</td>
<td></td>
<td>HPLC–ESI–MS</td>
<td>[180]</td>
</tr>
<tr>
<td>sinapic acid glucuronide</td>
<td>liver</td>
<td>C_{17}H_{20}O_{11}</td>
<td>400.1006</td>
<td>223, 208, 179, 164</td>
<td></td>
<td>HPLC–ESI–IT</td>
<td>[178]</td>
</tr>
<tr>
<td>trans-(p)-feruloyl alcohol</td>
<td>WG</td>
<td>C_{20}H_{20}O_{10}</td>
<td>428.1682</td>
<td></td>
<td></td>
<td>HPLC–LTQ–Orbitrap</td>
<td>[145]</td>
</tr>
</tbody>
</table>

### 2.2.8 Phytosterols

Phytosterols are plant-produced steroid compounds and triterpenes. They can be classified into plant sterols and stanols, the latter having a completely saturated sterol ring structure (Figure 2.10). Phytosterols exist in plants in their free forms and as esters; especially in cereals, they can be esterified with hydroxycinnamic acids [181]. Vegetable oils are the most abundant source of phytosterols in human diets; whole grains are also considered as an important source, with rye containing the highest amount (955 µg/g FW), followed by barley (761 µg/g FW), wheat (690 µg/g FW), and oats (447 µg/g FW) [182]. More than 250 different phytosterols have been found in plants, most common being sitosterol, stigmasterol and campesterol [181]. Rice bran contains γ-oryzanol, which is a mixture of phytosterols esterified with ferulic acid [169]. Despite common belief, many plants contain also cholesterol, the main animal sterol [181, 183].

**Biological activity.** The LDL cholesterol-lowering effect of phytosterols was first observed in humans in 1953 by Pollak [184]. Further evidence has established the effect, and as a result, the European Food Safety Authority (EFSA) concluded that a daily intake of 1.5–2.4 g of phytosterols can reduce the blood cholesterol level by 7–10.5% and thus allowed a health claim to be attached to food products containing a sufficient amount of these compounds [185]. Despite the convincing evidence for the cholesterol-lowering effect, the mechanisms behind it are yet to be confirmed. Several hypotheses have been proposed and two of them have endured the findings thus far:
phytosterols most likely work by impairing the micellar solubilization of cholesterol in the intestine thus reducing its absorption and/or by stimulating transintestinal cholesterol excretion (TICE) [186]. While increased LDL is associated with increased risk of cardiovascular disease, no association has been established with the serum phytosterol levels and CVD risk [181]. Research is ongoing also in the potential preventive effect of phytosterols in breast, liver, lung, ovary, prostate, and stomach cancer, via several proposed mechanisms [187]. However, most of the studies have been conducted in vitro and on animals.

**Analysis.** Phytosterols are relatively volatile and non-polar, and therefore they are most often quantified with GC–MS [181]. Modern LC–MS methods are also well suited for the analysis, but care should be taken in selecting the extraction solvent to ensure the solubility of the maximal amount of phytosterol species. Lipid extraction methods using a mixture of chloroform and methanol are commonly used [181]. Atmospheric pressure chemical ionization (APCI) is more commonly used as the ionization technique compared to electrospray ionization (ESI) [181]. In GC–MS, Esche et al. reported characteristic fragments at $m/z$ 255 for sterols and $m/z$ 213 for their fatty acid esters, and $m/z$ 257 for stanols and $m/z$ 215 for their fatty acid esters [188]. Ferulic and $p$-coumaric acid esters produce negative fragment ions at $m/z$ 177 and 163, corresponding with demethylated feruloyl and $p$-coumaroyl moieties, respectively [189]. The phytosterols detected in whole grains are presented in Table 2.8 with their LC–MS characteristics.
Figure 2.10. Chemical structures of phytosterols detected in whole grains

Table 2.8. Mass spectral characteristics of phytosterols detected in whole grains. MM = monoisotopic neutral molecular mass in Daltons.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Origin</th>
<th>Mol. formula</th>
<th>MM (Da)</th>
<th>Fragments m/z (−)</th>
<th>Fragments m/z (+)</th>
<th>Detection method</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>cholesterol</td>
<td>WG</td>
<td>C_{27}H_{46}O</td>
<td>386.3549</td>
<td>368, 353, 213, 371, 255, 129</td>
<td></td>
<td>GC–MS</td>
<td>[188]</td>
</tr>
<tr>
<td>β-cholesterol</td>
<td>WG</td>
<td>C_{27}H_{46}O</td>
<td>388.3705</td>
<td>355, 445, 460</td>
<td></td>
<td>GC–MS</td>
<td>[190]</td>
</tr>
<tr>
<td>campestanol</td>
<td>WG</td>
<td>C_{28}H_{48}O</td>
<td>402.3862</td>
<td>255, 213, 394, 129, 379, 397</td>
<td></td>
<td>GC–MS</td>
<td>[188]</td>
</tr>
<tr>
<td>stigmasterol</td>
<td>WG</td>
<td>C_{29}H_{48}O</td>
<td>412.3705</td>
<td>315, 299, 287, 261, 257, 243, 229, 215, 203, 189, 175, 161, 149, 147, 135</td>
<td></td>
<td>UPLC–APCI-LTQ–Orbitrap</td>
<td>[191]</td>
</tr>
<tr>
<td>β-sitosterol</td>
<td>WG</td>
<td>C_{29}H_{48}O</td>
<td>414.3862</td>
<td>215, 401, 398, 257, 383, 129</td>
<td></td>
<td>GC–MS</td>
<td>[188]</td>
</tr>
<tr>
<td>sitostanol</td>
<td>WG</td>
<td>C_{29}H_{48}O</td>
<td>416.4018</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Compound</td>
<td>Origin</td>
<td>Mol. formula</td>
<td>MM</td>
<td>Fragments $m/z$ (−)</td>
<td>Fragments $m/z$ (+)</td>
<td>Detection method</td>
<td>Ref.</td>
</tr>
<tr>
<td>---------------------------</td>
<td>--------</td>
<td>--------------</td>
<td>--------------</td>
<td>---------------------</td>
<td>---------------------</td>
<td>---------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>campestanly p-coumarate</td>
<td>WG</td>
<td>C$<em>{33}$H$</em>{55}$O$_3$</td>
<td>548.4229 503, 163</td>
<td>236, 219, 191, 384, 369, 164, 605</td>
<td></td>
<td>HPLC–APCI–LTO–Orbitrap, GC–MS</td>
<td>[188, 189]</td>
</tr>
<tr>
<td>campestryl ferulate</td>
<td>WG</td>
<td>C$<em>{33}$H$</em>{55}$O$_4$</td>
<td>576.4179 560, 545, 177</td>
<td>249, 633, 382, 266, 367, 221, 194</td>
<td></td>
<td>HPLC–EI/MS</td>
<td>[188]</td>
</tr>
<tr>
<td>24-methylthahosterol ferulate</td>
<td>WG</td>
<td>C$<em>{33}$H$</em>{55}$O$_4$</td>
<td>576.4179</td>
<td>576</td>
<td></td>
<td>HPLC–EI/MS</td>
<td>[83]</td>
</tr>
<tr>
<td>24-methylenecholestanol ferulate</td>
<td>WG</td>
<td>C$<em>{33}$H$</em>{55}$O$_4$</td>
<td>576.4179</td>
<td>194, 177, 394, 298, 256, 449, 367, 561</td>
<td></td>
<td>HPLC–EI/MS</td>
<td>[83]</td>
</tr>
<tr>
<td>campestanly ferulate</td>
<td>WG</td>
<td>C$<em>{33}$H$</em>{55}$O$_4$</td>
<td>578.4335 562, 534, 177</td>
<td>249, 266, 221, 384, 369, 635, 194</td>
<td></td>
<td>HPLC–APCI–LTO–Orbitrap, GC–MS</td>
<td>[188, 189]</td>
</tr>
<tr>
<td>24-methylcholesterol ferulate</td>
<td>WG</td>
<td>C$<em>{33}$H$</em>{55}$O$_4$</td>
<td>578.4335</td>
<td>578</td>
<td></td>
<td>HPLC–EI/MS</td>
<td>[83]</td>
</tr>
<tr>
<td>β-sitosteryl ferulate</td>
<td>WG</td>
<td>C$<em>{33}$H$</em>{55}$O$_4$</td>
<td>590.4335 574, 559, 546, 530, 177</td>
<td>396, 249, 266, 381, 647, 221, 194</td>
<td></td>
<td>HPLC–APCI–LTO–Orbitrap, GC–MS</td>
<td>[188, 189]</td>
</tr>
<tr>
<td>schottenol ferulate</td>
<td>WG</td>
<td>C$<em>{33}$H$</em>{55}$O$_4$</td>
<td>590.4335</td>
<td>590</td>
<td></td>
<td>HPLC–EI/MS</td>
<td>[83]</td>
</tr>
<tr>
<td>sitostany p-coumarate</td>
<td>WG</td>
<td>C$<em>{33}$H$</em>{55}$O$_4$</td>
<td>592.4492 576, 548, 532, 177</td>
<td>249, 649, 398, 266, 383, 221, 194</td>
<td></td>
<td>HPLC–APCI–LTO–Orbitrap, GC–MS</td>
<td>[188, 189]</td>
</tr>
<tr>
<td>24-methylenecycloar- tanyl ferulate</td>
<td>WG</td>
<td>C$<em>{41}$H$</em>{50}$O$_4$</td>
<td>616.4492 600, 585, 572, 177</td>
<td>249, 407, 422, 673, 266, 221, 194</td>
<td></td>
<td>HPLC–EI/MS</td>
<td>[83]</td>
</tr>
<tr>
<td>campestanly-16:0</td>
<td>WG</td>
<td>C$<em>{44}$H$</em>{52}$O$_2$</td>
<td>638.6002</td>
<td>382, 383, 367, 81, 255, 213</td>
<td></td>
<td>GC–MS</td>
<td>[188]</td>
</tr>
<tr>
<td>campestanly-16:0</td>
<td>WG</td>
<td>C$<em>{44}$H$</em>{52}$O$_2$</td>
<td>640.6158</td>
<td>384, 215, 385, 369, 257, 81</td>
<td></td>
<td>GC–MS</td>
<td>[188]</td>
</tr>
<tr>
<td>stigmasterly-16:0</td>
<td>WG</td>
<td>C$<em>{45}$H$</em>{54}$O$_2$</td>
<td>650.6002</td>
<td>394, 395, 379, 255, 81, 213</td>
<td></td>
<td>GC–MS</td>
<td>[188]</td>
</tr>
<tr>
<td>sitosteryl-16:0</td>
<td>WG</td>
<td>C$<em>{45}$H$</em>{54}$O$_2$</td>
<td>652.6158</td>
<td>255, 396, 213, 381, 397, 81</td>
<td></td>
<td>GC–MS</td>
<td>[188]</td>
</tr>
<tr>
<td>sitostanyl-16:0</td>
<td>WG</td>
<td>C$<em>{45}$H$</em>{54}$O$_2$</td>
<td>654.6315</td>
<td>215, 398, 383, 81, 399, 257</td>
<td></td>
<td>GC–MS</td>
<td>[188]</td>
</tr>
<tr>
<td>campesteryl-18:2</td>
<td>WG</td>
<td>C$<em>{46}$H$</em>{56}$O$_2$</td>
<td>662.6002</td>
<td>382, 383, 81, 367, 213, 255, 213</td>
<td></td>
<td>GC–MS</td>
<td>[188]</td>
</tr>
<tr>
<td>campestanly-18:2</td>
<td>WG</td>
<td>C$<em>{46}$H$</em>{56}$O$_2$</td>
<td>664.6158</td>
<td>385, 384, 81, 369, 215, 257, 255</td>
<td></td>
<td>GC–MS</td>
<td>[188]</td>
</tr>
<tr>
<td>campestanly-18:1</td>
<td>WG</td>
<td>C$<em>{46}$H$</em>{56}$O$_2$</td>
<td>664.6158</td>
<td>382, 383, 367, 81, 255</td>
<td></td>
<td>GC–MS</td>
<td>[188]</td>
</tr>
<tr>
<td>campestanly-18:1</td>
<td>WG</td>
<td>C$<em>{46}$H$</em>{56}$O$_2$</td>
<td>666.6315</td>
<td>385, 384, 81, 215, 257, 369</td>
<td></td>
<td>GC–MS</td>
<td>[188]</td>
</tr>
<tr>
<td>sitosteryl-18:3</td>
<td>WG</td>
<td>C$<em>{47}$H$</em>{58}$O$_2$</td>
<td>674.6002</td>
<td>396, 397, 81, 255, 381, 213</td>
<td></td>
<td>GC–MS</td>
<td>[188]</td>
</tr>
<tr>
<td>stigmasterly-18:2</td>
<td>WG</td>
<td>C$<em>{47}$H$</em>{58}$O$_2$</td>
<td>674.6002</td>
<td>394, 395, 255, 81, 379, 213</td>
<td></td>
<td>GC–MS</td>
<td>[188]</td>
</tr>
<tr>
<td>sitosteryl-18:2</td>
<td>WG</td>
<td>C$<em>{47}$H$</em>{58}$O$_2$</td>
<td>676.6158</td>
<td>397, 81, 396, 213, 255, 381</td>
<td></td>
<td>GC–MS</td>
<td>[188]</td>
</tr>
<tr>
<td>stigmasterly-18:1</td>
<td>WG</td>
<td>C$<em>{47}$H$</em>{58}$O$_2$</td>
<td>676.6158</td>
<td>394, 395, 255, 81, 379, 213</td>
<td></td>
<td>GC–MS</td>
<td>[188]</td>
</tr>
<tr>
<td>sitostany p-coumarate</td>
<td>WG</td>
<td>C$<em>{47}$H$</em>{58}$O$_2$</td>
<td>678.6315</td>
<td>399, 81, 398, 215, 257, 383</td>
<td></td>
<td>GC–MS</td>
<td>[188]</td>
</tr>
<tr>
<td>sitostany p-coumarate</td>
<td>WG</td>
<td>C$<em>{47}$H$</em>{58}$O$_2$</td>
<td>678.6315</td>
<td>396, 397, 81, 381, 255, 213</td>
<td></td>
<td>GC–MS</td>
<td>[188]</td>
</tr>
<tr>
<td>sitostany p-coumarate</td>
<td>WG</td>
<td>C$<em>{47}$H$</em>{58}$O$_2$</td>
<td>680.6471</td>
<td>399, 398, 81, 215, 257, 383</td>
<td></td>
<td>GC–MS</td>
<td>[188]</td>
</tr>
</tbody>
</table>
2.2.9 Tocols

Tocols encompass eight lipophilic compounds: four tocopherols (α-, β-, γ-, and δ-tocopherol) and four tocotrienols (α-, β-, γ-, and δ-tocotrienol), which are also commonly referred to as vitamin E [192]. The general structure of tocols includes a polar chromanol ring connected to a lipophilic 16-carbon chain, and the specific compounds differentiate based on the number and position of methyl groups in the chromanol ring [192] (Figure 2.11). In plants, seeds generally contain relatively high amount of tocols [192]. Wheat germ is one of the most abundant sources of vitamin E, containing 212 µg/g DM (α-tocopherol equivalent), while whole-grain flours typically contain 12–16 µg/g DM [193]. In wheat, α- and β-tocopherols and -tocotrienols are the dominant tocols, with tocotrienols more abundant than tocopherols [194].

**Biological activity.** Tocols are referred to as vitamin E because their deficiency, occurring rather from abnormalities in their absorption and metabolism than from an insufficient diet, will cause symptoms including peripheral neuropathy (ataxia) and haemolytic anaemia [195]. Tocols are among some of the most important antioxidants. The unusually weak O–H bond in the chromanol ring allows them to donate a hydrogen to peroxyl radicals, thus preventing the oxidation of e.g. polyunsaturated fatty acids (PUFA) in cell membranes and lipoproteins in plasma [192, 196]. Peroxyl radicals react 1000 times faster with tocols compared to PUFAs [195]. The tocopheryl radical is reduced back to tocopherol by reacting with a hydrogen donor, such as vitamin C [195]. Interestingly, all the stereoisomers of tocols arising from the three chiral carbons at 2′, 4′, and 8′ positions have antioxidant activity, but only the RRR-configuration has significant biological activity, and it is the naturally occurring form of the compounds [195]. Since vitamin E was discovered in 1922, several potential health benefits have been associated with it, including prevention of lipidemia and hypertension, suppression of allergic dermatitis, protection of kidneys and the nervous system, and anti-inflammatory activities [192].

**Analysis.** Tocols are lipophilic and unstable molecules, degrading under the presence of UV light, high temperature and oxidizing agents, which creates challenges for their analysis [197]. Most semi-polar or non-polar organic solvents are suitable for the extraction of tocols [197]. Tocols do not possess polar groups capable of protonation, hindering their detectability particularly in ESI-MS [197]. Adding formic acid or ammonium or metal ions in the mobile phase enhances the protonation and enables the detection of the compounds in samples where they are present in relatively high concentration [192, 198]. Using C30 as the reversed-phase HPLC column can also enhance the separation of these compounds [199]. Tocols can be detected in either the positive or negative ionization mode, producing [M + H]+ and [M – H]− ions, respectively [192, 197]. According to Lampi et al., α-tocols produce characteristic fragment ions at m/z 205 and 165 and β- and γ-tocols at m/z 191 and 151 in LC–MS with atmospheric pressure chemical ionization (APCI) [194]. Table 2.9 reviews the tocol species in whole grains with their observed LC–MS characteristics.
Table 2.9. Mass spectral characteristics of tocols detected in whole grains. MM = monoisotopic neutral molecular mass in Daltons.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Origin</th>
<th>Mol. formula</th>
<th>MM</th>
<th>Fragments m/z (−)</th>
<th>Fragments m/z (+)</th>
<th>Detection method</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>β-tocotrienol</td>
<td>WG</td>
<td>C_{28}H_{43}O₂</td>
<td>410.3185</td>
<td>149, 394, 339</td>
<td>191, 151</td>
<td>HPLC–ESI–MS, HPLC–DAD–MS</td>
<td>[194, 200]</td>
</tr>
<tr>
<td>γ-tocotrienol</td>
<td>WG</td>
<td>C_{28}H_{43}O₂</td>
<td>410.3185</td>
<td>177, 137</td>
<td></td>
<td>HPLC–ESI–MS</td>
<td>[198]</td>
</tr>
<tr>
<td>β-tocopherol</td>
<td>WG</td>
<td>C_{28}H_{42}O₂</td>
<td>416.3654</td>
<td>400, 149</td>
<td>191, 151</td>
<td>HPLC–ESI–MS, HPLC–DAD–MS</td>
<td>[194, 200]</td>
</tr>
<tr>
<td>γ-tocopherol</td>
<td>WG</td>
<td>C_{28}H_{42}O₂</td>
<td>416.3654</td>
<td>177, 137</td>
<td></td>
<td>HPLC–ESI–MS</td>
<td>[198]</td>
</tr>
<tr>
<td>α-tocotrienol</td>
<td>WG</td>
<td>C_{20}H_{35}O₂</td>
<td>424.3341</td>
<td>205, 165</td>
<td></td>
<td>HPLC–ESI–MS</td>
<td>[198]</td>
</tr>
<tr>
<td>α-tocopherol</td>
<td>WG</td>
<td>C_{20}H_{35}O₂</td>
<td>430.3811</td>
<td>163, 414</td>
<td>205, 165</td>
<td>HPLC–ESI–MS, HPLC–DAD–MS</td>
<td>[198]</td>
</tr>
</tbody>
</table>

2.3 EFFECT OF FOOD PROCESSING ON WHOLE GRAIN PHYTOCHEMICALS

2.3.1 Sourdough

Sourdough is a fermentation process of cereal products, where the dough is incubated with yeast and lactic acid bacteria (LAB). It likely dates to ancient Egypt, where probably appeared spontaneously by microbes already present in the grains; it was used in beer brewing and bread baking [201]. Typical genera of LAB used in sour-
dough include Enterococcus, Lactobacillus, Leuconostoc, and Weissella [202]; typical species are Lactobacillus brevis and Lactobacillus plantarum, accompanied with yeast species, such as baker’s yeast (Saccharomyces cerevisiae) and Candida milleri [56]. Sourdough has particular importance in the leavening of whole-grain rye bread, where it improves the technical properties of the bread, including texture, flavour and shelf-life [203]. It is also used in certain artisanal wheat breads, such as the San Francisco bread [56]. Sourdough starters with special metabolic properties have been used to decrease the FODMAP (fermentable oligosaccharides, disaccharides, monosaccharides, and polyols) content of rye bread to reduce the symptoms caused by these compounds in people affected with irritable bowel syndrome (IBS) [204].

In addition to technical improvements, sourdough has potential in increasing the health-promoting properties of bread: it allows the inclusion of more bran in the bread and it also modifies the bran section, releasing phytochemicals from the fibre matrix with e.g. feruloyl esterase, an enzyme hydrolysing the bonds between ferulic acid and arabinoxylan [203, 205-207]. We have recently reviewed the effect of sourdough on specific classes of phytochemicals [8] and a more comprehensive review on the conversions of nutrients and bioactives has been published by Gänzle [208]. The knowledge is briefly summarized and updated herein. There is some evidence of the degradation of alkylresorcinols by sourdough fermentation [74, 209]; however, other studies have shown no significant effect [206] or even a moderate increase [56], and the increase of known alkylresorcinol metabolites (3,5-dihydroxybenzoic acid and 3(3,5-dihydroxyphenyl)propanoic acid) has not been reported. Regarding benzoazinoids, sourdough microbes are capable of hydrolysing benzoazinoid glucosides and possibly metabolizing the aglycone forms even further into 1,3-benzoazin-2-one (BOA) and N-(2-hydroxyphenyl)acedamide (HPAA) [101, 102, 210]. Other groups of phytochemicals affected by sourdough fermentation include folates (increased levels) [211], phenolic acids (transformations into metabolites) [208, 212] and tocols (decreased levels) [206, 213]. We have recently showed that sourdough induces major changes in the metabolite profile of sourdough bread: it increases the abundance of branched-chain amino acids (BCAA) and their metabolites and produces small peptides and phenolic acid metabolites (dihydroferulic and dihydrocaffeic acid), which have potential antioxidative and other health effects [56].

2.3.2 Enzymatic bioprocessing

Compared to sourdough, bioprocessing with cell wall degrading enzymes is a more recent and technological means to improve the nutritional and technical quality of whole-grain products. The outer layers of cereal grains naturally contain such enzymes, such as endoxylanases, the purpose of which is to hydrolyse the endosperm cell wall and make the starch and proteins (mainly gluten) within the endosperm accessible during germination [214]. Using the enzymes as a pre-treatment before fermentation can improve the sensory properties and the texture and stability of breads supplemented with wheat bran [215]. Mateo Anson et al. treated wheat bran with a mixture of enzymes (xylanase, β-glucanase, α-amylase, cellulase, and ferulic
acid esterase) and observed a 5-fold increase in the bioaccessibility of ferulic acid, the most abundant phenolic compound in rye (from 1.1% to 5.5%) and increased levels of phenolic acid metabolites (mainly 3-(3-hydroxyphenyl)propionic acid and 3-phenylpropanoic acid) in an in vitro model of the human gastrointestinal system [216]. In rye bran, which was added to white wheat bread bioprocessing with a mixture of hydrolytic enzymes increased the bioaccessibility of several phenolic acids (5,5’-diferulic acid, ferulic acid, p-coumaric acid, sinapic acid) but also benzoxazinoids (BOA, HMBOA, HMBOA glycoside) and flavonoids (luteolin, apigenin 7-O-glucuronide, isoorientin) [74], which may have been released from their corresponding mono- or diglycosides by enzymatic activity. The changes in the phytochemical profile induced by bioprocessing were different to those caused by sourdough fermentation of whole-grain rye bread [74]. As the bioavailability of phytochemicals can be a major bottleneck in their health-promoting effects, bioprocessing offers a promising technological approach to further increase the nutritional value of whole-grain products.

2.4 GUT MICROBIAL METABOLISM OF WHOLE GRAIN PHYTOCHEMICALS

Humans and other mammals have a long history of co-evolution with symbiotic microbial communities inhabiting the gut [217, 218]. While the bacteria, archaea and fungi living in the gut are provided with a unique habitat and constant supply of food, the also benefit the host by extracting additional nutrients and bioactive compounds from the food, maintaining a healthy immune system and resisting the invasion of pathogens [219]. Compared to their hosts, gut microbial communities possess a far wider range of enzymes capable of performing a multitude of catalytic reactions on phytochemicals: these include hydrolysis of sugar moieties, decarboxylation, deconjugation, dehydroxylation, demethylation, oxidation, reduction, racemisation and ring fission [40, 220-222]. There is great interindividual variability in the composition of microbiota, and individuals can be divided into different enterotypes based on the dominant bacterial genus (mainly Bacteroides and Prevotella) [223]. Most types of microbial communities can metabolise a similar variety of phytochemicals, but there are exceptions: isoflavones daidzein and genistein are metabolised into equol much more efficiently in about one third of the population, with over 600-fold difference in equol production between the high and low producers [224]. Certain bacterial strains, including Eggerthella sinensis and Adlercreutzia equolifaciens, have thus far been associated with equol metabolism [225]. Intraindividual variability in the metabolism may occur as well; diet heavily influences the gut microbial composition [226, 227] and a diet rich in cereal bran also has an impact of its own [41, 42]. Orally administered antibiotics, usually targeted against pathogenic bacteria elsewhere in the body, alter the gut microbiota and may inhibit the production of enterolactone from dietary lignans, a compound that is produced exclusively by gut microbes [228].
The transformation of phytochemicals in the colon by gut microbes has three main potential effects: 1) the conversion often increases the bioavailability of phytochemicals by producing metabolites that are smaller in size and thus are absorbed better into circulation; 2) the microbial metabolites have health benefits of their own; and 3) dietary phenolic compounds can also modulate the composition of gut microbiota in a beneficial way by suppressing the growth of potential pathogens, such as *Clostridium difficile* and *Escherichia coli* [40, 229]. Microbial metabolism of phytochemicals present in whole grains (and other plant-based foods) thus plays an important part in mediating the health effects of diets rich in such foods; it may be the metabolites that are rather responsible for the health benefits rather than the original compounds in plants, which may not reach their target in their intact form.

We have recently reviewed the microbial metabolism of phytochemicals contained in rye [8]. The current knowledge is briefly reviewed and updated herein.

### 2.4.1 Microbial metabolism of whole grain phytochemicals

**Benzoxazinoids.** Currently, the potential microbial metabolism of benzoxazinoids is under investigation and further research is needed to confirm the extent to which gut microbiota is contributing to the metabolism. Benzoxazinoid glycosides, such as HBOA hexose, may be deconjugated and the aglycone metabolised into uncharacterised compounds in the colon, as suggested by an *in vitro* experiment by Hanhineva et al. [79]. Microbes in the soil transform benzoxazinoids via 2-aminophenol into phenoxazinones, such as 2-aminophenoxazin-3-one (APO) and 2-acetylaminophenoxazin-3-one (AAPO) [230]. These metabolites have been detected from pigs after rye-enriched diet [100] but not from humans [231]. The potential site of metabolism for phenoxazinone metabolites remains unclear.

**Betaines.** Dietary betaine is rapidly absorbed and distributed in the circulation and metabolised into dimethylglycine in the liver and kidney [232]. It has been hypothesised that gut microbiota may transform glycine betaine into other amino acid-derived betaines [80] if the compound reaches the colon using e.g. fibre matrix as a carrying vehicle. Pekkinen et al. reported several amino acid-derived betaines, including alanine betaine, phenylalanine betaine, pipecolic acid betaine, proline betaine, trigonelline, and valine betaine, from the urine of mice fed with a rye bran-supplemented diet [80]. Glycine betaine, the most abundant betaine in whole grains, was not present in the urine, suggesting that it may have been metabolised into the observed compounds by e.g. colonic microbes. We recently observed that a bran-enriched feed elevates the same betainized compounds, and additionally 5-aminovaleric acid betaine and tryptophan betaine, in the colonic content of mice [43]. In germ-free mice, the levels of these compounds were significantly lower compared to conventional mice, indicating the contribution of microbiota in betaine metabolism. The results were further confirmed in an *in vitro* model of the human gastrointestinal system incubated with rye bran, where the production of 5-aminovaleric acid betaine, alanine betaine, proline betaine, and trigonelline continued throughout the ferme-
tation by colonic microbes whereas the levels of glycine betaine were reduced to faecal background at the end of the incubation. While positive associations were observed between the relative levels of betaines and *Bifidobacterium* in the caecal contents of mice, the mechanism and bacterial species or taxa behind the metabolism remains to be elucidated.

**Lignans.** The conversion of plant lignans into mammalian lignans (enterolignans), namely enterodiol and enterolactone, is one of the most well-studied examples of colonic microbial metabolism of phytochemicals, first reported in 1982 by Axelson et al. [233]. In whole-grain cereals, the known precursors of enterolignans include lariciresinol, matairesinol, pinoresinol, secoisolariciresinol, secoisolariciresinol diglucoside, and syringaresinol [234, 235]; other suggested but unconfirmed precursors are medioresinol and buddlenol- and hedyotisol-type sesquilignans and dilignans [165]. Despite the fact that enterolignan metabolism requires a relatively complex series of catalytic reactions, it is highly prevalent in humans, which may be explained by the widespread and diverse microbial communities capable of performing the complete set of reactions [236-238]. The rate of metabolism and the end product depends on the precursor: for instance, matairesinol is transformed primarily into enterolactone, and only a small fraction of syringaresinol is metabolised by the gut microbiota [235]. Overall, the microbial metabolism of plant lignans has important health implications because it can increase the bioavailability [239], antioxidative effects [162], and oestrogen activity [156, 157] of lignans obtained from the diet.

**Phenolic acids and flavonoids.** The free phenolic acids are likely absorbed already in the small intestine, based on the rapid (1–3 hours) increase of ferulic acid in plasma after rye intake [240, 241] However, the portion of freely soluble phenolic acids in whole grains is very low, up to 1% [242], and thus a great majority of the compounds will reach the colon within the insoluble fibre matrix. Booth and Williams first reported in 1963 the colonic microbial metabolism of orally ingested caffeic acid in rat and rabbit caecal contents [243]. Kern et al. observed low levels of hydroxycinnamic acid precursors – the most abundant phenolic acids in whole grains – in human plasma after 6 hours from the rye intake, which is the expected time when the compounds would be released in the colon by microbial fermentation of the fibre [240]. This suggested that once released, they are for the most part transformed into metabolites prior to their absorption. It was shown that colonic microbes are able to release hydroxycinnamic acids and their dimers from the bran fibre matrix [244-246] and that these compounds are rapidly transformed into metabolites in vitro [176, 177, 247]. The routes of colonic microbial metabolism of phenolic acids are interconnected: ferulic acid may be a precursor to other hydroxycinnamic acids present in cereals [177], and to further complicate the elucidation of the metabolic routes, phenolic acids themselves may be resulting from ring fission of flavonoids by gut microbiota, resulting in the same end products [221]. In cereals, phenolic acids are more abundant compared to flavonoids, which suggests that the metabolites are mainly
originating from phenolic acid precursors. The main microbial metabolites of phenolic acids include 3,4-dihydroxyphenylpropionic acid (dihydrocaffeic acid), dihydroferulic acid, 3-phenylpropionic acid, 3-(3-hydroxyphenyl)propionic acid, and 3-coumaric acid (Table 2.7, Figure 7.5) [177, 216, 243, 247]. Braune et al. determined in vitro the metabolic pathway of 8-O-4-diferulic acid into certain main metabolites, which included homovanillic acid, 3-(3,4-dihydroxyphenyl)propionic acid, and 3,4-dihydroxyphenylacetic acid [247]. While the health effects of phenolic acids and their mechanisms are not yet established, it is likely that they are largely attributable to the microbial metabolites due to the extensive metabolism in the gut.

2.5 MASS SPECTROMETRY-BASED NONTARGETED ANALYSIS OF PHYTOCHEMICALS

Plants are the master chemists of living organisms: the total number of different metabolites produced by them has been estimated as high as 1 060 000 [248] and even in an individual plant, the number can reach more than a thousand [249]. This poses a challenge for nontargeted metabolomics, which aims for a global-scale detection and identification of metabolites. Indeed, no single method is currently capable of performing such task completely, but several major phytochemical groups, including flavonoids and phenolic acids, are well within reach of typical MS instruments [250]. Mass spectrometry is the most widely used method in the analysis of phytochemicals due to its high sensitivity (detectable concentrations down to picomolar range), selectivity, dynamic range, and accuracy [251-254]. A high mass accuracy of modern mass spectrometers, reaching 1 ppm or less, can allow a limited level of structural elucidation. Nuclear magnetic resonance (NMR) spectroscopy is less suitable for analysing a large number of metabolites with low concentrations due to its lower sensitivity compared to mass spectrometry; however, it offers some major advantages, such as high-throughput analysis, easy sample preparation, straightforward quantification, and full structure elucidation of novel compounds [11]. The analytical methods used in the analysis of specific classes of phytochemicals are reviewed in Chapter 2.2 and summarised in Table 2.10.
<table>
<thead>
<tr>
<th>Compound class</th>
<th>Mass range* [Da]</th>
<th>Extraction solvent</th>
<th>Chromatography</th>
<th>Ionization mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alkylresorcinols</td>
<td>318–578</td>
<td>acetone, ethyl acetate, (methanol)</td>
<td>GC, LC (RP)</td>
<td>+ –</td>
</tr>
<tr>
<td>Benzoazinoids</td>
<td>135–859</td>
<td>methanol</td>
<td>LC</td>
<td>(+) –</td>
</tr>
<tr>
<td>Betaines</td>
<td>117–246</td>
<td>water, methanol</td>
<td>LC (HILIC)</td>
<td>+</td>
</tr>
<tr>
<td>Carotenoids</td>
<td>536–568</td>
<td>acetonitrile, methanol</td>
<td>LC (e.g. RP C30)</td>
<td>+</td>
</tr>
<tr>
<td>Flavonoids</td>
<td>270–898</td>
<td>aqueous methanol or acetonitrile</td>
<td>LC</td>
<td>+ (anthocyanidins: +)</td>
</tr>
<tr>
<td>Lignans</td>
<td>298–840</td>
<td>non-polar (e.g. hexane) + polar (methanol)</td>
<td>GC, LC</td>
<td>+ –</td>
</tr>
<tr>
<td>Phenolic acids</td>
<td>122–428</td>
<td>aqueous methanol or acetone</td>
<td>LC</td>
<td>+ –</td>
</tr>
<tr>
<td>Phytosterols</td>
<td>386–680</td>
<td>chloroform:methanol</td>
<td>GC, LC</td>
<td>+ (–)</td>
</tr>
<tr>
<td>Tocots</td>
<td>410–430</td>
<td>most polar and semi-polar solvents</td>
<td>LC</td>
<td>+ –</td>
</tr>
</tbody>
</table>

### 2.5.1 Chromatographic separation

To facilitate the separation of compounds in MS, the detection is usually performed after a chromatographic separation technique, such as liquid chromatography (LC) or its more advanced forms HPLC (high-performance liquid chromatography) and UHPLC (ultra-high-performance liquid chromatography), gas chromatography (GC), or high-performance thin-layer chromatography (HPTLC) [255]. Ion mobility (IM) is a relatively new technique, which can be incorporated between the chromatography and MS, adding another dimension to the separation of the compounds by measuring the collision cross section (CCS) of each analyte [256]. Supercritical fluid chromatography (SFC) is used in the separation of lipophilic compounds due to its high resolution and modifiable polarity of the mobile phase [124]. Another separation method resembling chromatography is capillary electrophoresis (CE), which is based on the electrophoretic mobility depending on the charge and size of the analytes [257]. Out of these, UHPLC–MS is the prevailing method in analysing phytochemicals from various matrices, such as food and body fluids [250, 253, 254, 258]. While GC–MS offers high reproducibility and the possibility to use retention time index (RTI) to assist with metabolite identification, it has a limited mass range and requires derivatization of non-volatile compounds [250, 253].

In LC–MS, several types of columns are available for the separation of molecules. The most common in (plant) metabolomics is the reversed-phase (RP) column, having hydrophobic alkyl chains (e.g. C18) bonded to the silica particles, together forming the stationary phase [259]. Therefore, lipophilic compounds have a longer interaction...
with the stationary phase and elute later into the mass spectrometer with good separation, while extremely polar compounds have virtually no interaction and elute rapidly in the very beginning of the chromatography with poor separation. Polar compounds can be better separated with hydrophilic interaction chromatography (HILIC) and porous graphitic carbon (PGC) columns, which have an orthogonal principle compared with RP, having e.g. hydrophilic zwitterions attached to the silica particles (HILIC) or graphite (PGC) as the stationary phase [259]. As in the sample preparation step, where two extraction methods can be combined to maximise the coverage of metabolites, reversed-phase chromatography can be complemented with an analysis using HILIC or PGC column to allow the detection of both extremes of the polarity range.

2.5.2 Applications of mass spectrometry

Ionisation of molecules is a prerequisite for detection in mass spectrometry because the technology is based on the measurement of the mass-to-charge ratio (m/z) of ions. Several techniques exist for the ionisation and they can be classified based on the “softness” of the ionisation and whether it occurs in a vacuum or in atmospheric pressure [260-262]. Electron ionisation (EI), one of the earliest methods, is a hard ionisation technique, i.e. it produces a positive radical cation (M+●), which is often so extensively fragmented that it does not leave a trace of itself [261]. This can, however, be used as an advantage especially in GC–MS for the identification of the compounds. Fast atom bombardment (FAB), thermospray ionisation (TSI), and atmospheric pressure ionisation (API) are soft ionisation techniques, which typically produce protonated or deprotonated molecular ions ([M + H]+ or [M – H]–) or cation/anion adducts of polar neutral molecules (e.g. [M + Na]+ or [M + Cl]–) [261]. In API, the main ionisation sources used are electrospray ionisation (ESI), atmospheric pressure chemical ionisation (APCI), and atmospheric pressure photoionisation (APPI), which are well suited for coupling with liquid chromatography [260]. ESI is based on applying a high voltage to the end of a capillary where the liquid sample is ejected, causing it to be dispersed into increasingly small charged droplets until single ions are remaining [263]. In APCI, the ionization occurs in the gaseous phase and in APPI, which is basically another application of APCI, by photons from an ultraviolet lamp [7]. Soft ionization techniques require an additional fragmentation step by e.g. collision-induced dissociation (CID) to produce the characteristic fragmentation spectra (MS/MS or MS2 spectra) used in the identification of the metabolites [261]. From the perspective of nontargeted analysis of phytochemicals, the analytical range of each ionization technique is important to take into consideration. Thermospray is suitable for moderately polar compounds with a mass of 200 to 800 Da [260], excluding many small phenolic compounds, such as phenolic acids. ESI and APCI have a wider mass range, ESI being more suited for semi-polar and polar compounds (e.g. most polyphenols) and APCI for less polar and even neutral compounds, such as plant sterols [260, 264]. Since APCI requires volatilisation of the analytes, it is less suitable for thermolabile phytochemicals, including anthocyanidins [261]. Currently, ESI is probably
the predominant ionisation method used in nontargeted phytochemical analyses [76].

The mass analyser measures the mass-to-charge ratio of the ionised compounds. The most simple such instrument is the single quadrupole, which has major limitations in its resolution but which can be used to detect masses up to 4000 Da, including large tannins [265]. Tandem mass spectrometry includes two or more stages, where the ions are selected, fragmented, reselected and measured [262]; the selection of ions can be accomplished by separating them in space (triple quadrupole), in time (time-of-flight, TOF), or in both space and time (quadrupole time-of-flight, QTOF). QTOF offers a nearly unlimited mass range combined with high sensitivity and accuracy [265], making it suitable for nontargeted analyses, where rapid and accurate analysis of a wide range of compounds is required [76]. Other instruments with similar benefits include Fourier transform ion cyclotron resonance (FT-ICR) and linear trap quadrupole (LTQ) Orbitrap, the development of which has further increased the capacity, accuracy and resolution (the ability to separate mass peaks) of mass spectrometry, increasing the popularity of MS in nontargeted metabolomics.

2.5.3 Sample preparation

In non-targeted metabolomics, sample uniformity is crucial to avoid introducing bias when determining the differential metabolites, which should only originate from the biological processes related to the study setting, not from the analytical procedures themselves. Sample preparation in nontargeted plant metabolomics is a fairly simple process, including the homogenization of frozen (~80°C) material and the extraction of metabolites with an appropriate solvent [259]. In an ideal case, the solvent should extract all the metabolites present in the sample, it should not introduce changes in the metabolite levels, and it should not chemically modify the metabolites. Polar and semi-polar metabolites can be extracted with methanol or 80% aqueous methanol, which still allows the detection of many lipophilic compounds, such as alkylresorcinols [74]. For lipidomics and the analysis of mainly lipophilic phytochemicals, such as carotenoids and tocols, a non-polar solvent, such as chloroform, is more suitable. To maximise the metabolite yield in the solvent, a two-phase solvent system containing a mixture of chloroform, methanol and water can be used, and it fractionates the metabolites into an aqueous and organic phase, which can be analysed separately [259].

2.5.4 Data analysis

Modern LC–MS metabolomics platforms can detect more than 7000 distinct molecular features (ions with a unique combination of mass-to-charge ratio and retention time) from a single biological sample [266]. Since often the sample is analysed several times in the positive and negative ionization mode in one or more column types (e.g. reversed-phase and HILIC), the amount of data is in the gigabyte range and far beyond manual curation. Raw signals from an LC–MS instrument can be acquired in
centroid or profile mode. Centroid data reduces all the signals into a single data point with a certain $m/z$ and intensity; this also reduces the data size considerably but comes with drawbacks, such as loss of intricate isotopic information that could be used for determination of the molecular formula [267]. Once the raw data is obtained, it can be processed with vendor software or transformed into an open-source format (e.g. mzXML, mzData or mzML) for processing with a third-party or open-source software. Due to restrictions in computational capacity, the raw data is usually first compressed by e.g. binning, where the data is transformed into a matrix representation [267]. Since the instrument may incorporate systematic bias into the measurements, the next step is to normalise the data by e.g. using mathematical models or quality control (QC) samples (prepared by pooling from all the study samples), which are analysed regularly during the analytical run. Whereas normalisation is used to correct the data within chromatograms, scaling and transformation are orthogonal corrections performed within samples to allow semi-quantitative comparison of metabolite levels [267]. Examples of scaling algorithms include autoscaling [268] and Pareto scaling [269], and transformation is the nonlinear conversion of the data, e.g. into logarithmic scale.

Peak picking or feature detection is the process of collecting the molecular features from the pre-processed data (ions with certain $m/z$ and retention time), and importantly, separating them from the inherent background noise and false peaks the instrument has produced. Several algorithms have been developed for the peak picking, including continuous wavelet transformation (CWT) of mass spectra to obtain more information about the peak shapes to identify low intensity peaks from the background noise [270]. To compare peaks across samples and runs, peak alignment is performed to match the peaks originating from the same metabolites but having slightly variable retention time in each sample [267].

Finally, before the identification of the metabolites, it is necessary to determine the differential metabolites (or biomarkers) between the sample groups, which are relevant for the research question – including the whole dataset for identification is usually not sensible or even feasible. Widely used statistical models to define the differential metabolites include some classical hypothesis tests, such as $t$-test, Mann–Whitney U test, and analysis of variance (ANOVA), and modern chemometric methods, including principal component analysis (PCA), partial least squares discriminant analysis (PLS-DA), and variables importance of projection (VIP) [267]. In metabolomics data, the number of variables (metabolites) usually far exceeds the number of samples, which may lead to increased false discovery rate in multiple hypothesis testing [267]. For example, if one sets the significance threshold of the raw $p$-value acquired from a hypothesis test at commonly used 0.05, it is easy to find several “biomarkers” that are in fact completely sporadic in a typical metabolomics dataset, which contains hundreds or thousands of $p$-values (for each metabolite). As reviewed by Broadhurst and Kell, a more realistic approach is to use a more stringent threshold, such as the Bonferroni correction, where the required $p$-value for statistical significance is calculated by dividing the chosen threshold (say 0.05) by the number of
metabolites being tested for significance [271]. In case of a dataset containing 5000 metabolites, the p-value threshold for statistical significance would be $1 \times 10^{-5}$.

### 2.5.5 Metabolite identification

In non-targeted metabolomics, metabolite identification is a key process before the interpretation of the results. In contrast to NMR, complete elucidation of novel chemical structures is not possible in mass spectrometry, and thus the identification process in MS-based non-targeted metabolomics relies on existing identifications made with reference standards. GC–MS and LC–MS instruments produce several types of data that can be compared with existing information to deduce the identity of the molecular feature, and a single chemical parameter should not be used alone for the identification [272, 273]. The exact mass (i.e. the measured mass-to-charge ratio, $m/z$) of the molecular feature can be used in making queries in compound databases, provided that the mass accuracy is sufficient (< 3 ppm) to differentiate between similar masses (e.g. 176.0321 Da of glucuronide and 176.0473 Da of feruloyl group) and to determine the molecular formula [274, 275]. The molecular formula itself can be predicted based on certain rules, including the restricted number of elements, isotopic pattern (all isotopic combinations of the molecular feature produce separate peaks in mass spectrometry), the ratio of hydrogen and carbon and other elements, and the probabilities of these ratios [274]. Retention time from liquid or gas chromatography provides information about the lipophilicity or volatility, respectively, and can be compared with in-house libraries (LC–MS) and retention time index databases (GC–MS). In case a reference standard is not available for reliable identification, which is often the case with phytochemicals, such as flavonoid glycosides, it is essential to produce MS/MS spectra to obtain a characteristic pattern (or “fingerprint”) of the fragment ions. These can be produced by e.g. collision-induced dissociation (CID) in LC–MS. The MS/MS spectra can be compared with publicly available spectral libraries (Table 2.11) to putatively annotate the metabolite with high (but not 100%) reliability.
Table 2.11. The main publicly available mass spectral libraries, the methods used to acquire reference spectra, and the number of mass spectra or individual compounds included in the databases based on the information available in November 2018.

<table>
<thead>
<tr>
<th>Database</th>
<th>Method(s)</th>
<th>Number of mass spectra</th>
</tr>
</thead>
<tbody>
<tr>
<td>FiehnLib</td>
<td>GC–MS</td>
<td>Experimental 2 212</td>
</tr>
<tr>
<td></td>
<td></td>
<td>in silico 0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total 2 212</td>
</tr>
<tr>
<td>Golm Metabolome Database</td>
<td>GC–MS</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&gt; 26 500</td>
</tr>
<tr>
<td>Human Metabolome Database (HMDB)</td>
<td>GC–MS, LC–MS</td>
<td>29 665^a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>318 249^a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>347 914^a</td>
</tr>
<tr>
<td>LIPID MAPS and LipidBlast</td>
<td>LC–MS</td>
<td>21 312^b</td>
</tr>
<tr>
<td></td>
<td></td>
<td>21 953^b</td>
</tr>
<tr>
<td></td>
<td></td>
<td>43 265^b</td>
</tr>
<tr>
<td>MassBank (Japan)</td>
<td>GC–MS, LC–MS</td>
<td>50 998^a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50 998^a</td>
</tr>
<tr>
<td>METLIN</td>
<td>LC–QTOF-MS</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&gt; 1 000 000^b</td>
</tr>
<tr>
<td>MoNA (MassBank of North America)</td>
<td>GC–MS, LC–MS</td>
<td>120 429^a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>139 746^a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>260 175^a</td>
</tr>
<tr>
<td>mzCloud</td>
<td>LC–MS</td>
<td>8 384^b</td>
</tr>
<tr>
<td></td>
<td></td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 887 909^a</td>
</tr>
<tr>
<td>NIST 17</td>
<td>GC–MS, LC–MS</td>
<td>665 070^a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>665 070^a</td>
</tr>
</tbody>
</table>

Since metabolite identification determines how the results of a nontargeted study can be interpreted, the reliability of the identifications should be carefully assessed, and the relevant metadata for the identifications (such as characteristic MS/MS fragments) needs to be presented [272]. According to the proposal by Sumner et al., four levels of metabolite identifications can be determined based on the supporting evidence: 1. identified compounds based on comparison with an authentic standard analysed in the same platform; 2. putative annotations without a standard, based on spectral similarity to an entry in a relevant spectral library; 3. putative characterization of a compound class based on physicochemical or spectral similarity to a chemical class; and 4. unknown compounds, for which only the LC–MS characteristics can be reported [272]. The level of identification is not important only for the reliability of the reported metabolites; it also determines whether they need to be confirmed with reference standard for any future targeted studies.
3 AIMS OF THE STUDY

The aim of this doctoral thesis was to study the phytochemicals in whole grains and how they are transformed by food processing techniques, mainly enzymatic bioprocessing and sourdough fermentation. The other main aim was to investigate the microbial metabolism of whole grain phytochemicals by gut microbiota in an in vitro model of the human gut and in animal models.

The specific aims of the doctoral thesis were:

1. to compare the phytochemical profiles of traditional sourdough-fermented rye bread, white wheat bread fortified with intact rye bran, and white wheat bread fortified with enzymatically bioprocessed rye bran, using regular white wheat bread as a control, and to increase the number of characterised phytochemicals in rye (Study I);
2. to compare the metabolite profiles of sourdough-fermented whole-grain rye and wheat breads and to study the impact of sourdough on the potential health benefits of whole-grain products (Study II);
3. to study the influence of bran bioprocessing on the metabolism of phenolic acids and elucidating potential metabolic routes of ferulic acid in an in vitro model of the human gut and by utilizing GC × GC–TOF-MS and UPLC–QTOF-MS (Study III); and
4. to study the impact of bran-enriched diets on the gut microbial composition of C57BL/6J mice and to determine the potential contribution of gut microbiota in the metabolism of glycine betaine into other betainized compounds (Study IV).
Sample preparation. The samples analysed for the studies included in this dissertation were wheat bread fortified with unprocessed and bioprocessed rye bran (Study I), sourdough-fermented whole-grain rye bread (Study I and II), flours, doughs, and breads prepared from endosperm and whole-grain wheat and rye with and without sourdough (Study II), unprocessed and bioprocessed rye bran fermented in the in vitro colon model (Study III and IV), caecal and colonic contents of C57BL/6J mice fed with bran-enriched diets (Study IV), and samples of 13 different tissues (plasma, heart, liver, pancreas, muscle, duodenum, jejunum, ileum, cecum, colon, visceral adipose tissue, subcutaneous adipose tissue, and brown adipose tissue) from conventional murine pathogen free and germ-free C57BL/6NTac mice. Figure 4.1 gives an overview of the nontargeted metabolomics workflow used in the studies.

The samples were prepared by frozen tissue homogenization (if necessary) and adding 80% v/v (Studies I–III) or 90% v/v (Study IV) aqueous HPLC grade methanol in a ratio of 300 µl per 100 mg of sample (500 µl per 100 mg for colon contents in Study IV) for the metabolite extraction and protein precipitation. In Study II, formic acid (0.1% v/v) was added in the extraction solvent. The samples were subsequently vortexed and centrifuged (generally 10 min at 4 °C and 13 000 rpm), and the supernatant was collected and filtered into HPLC vials and kept in cold until analysis.

In vitro digestion and colon model. The in vitro fermentation of rye bran was performed as described by Nordlund et al. [285]. Briefly, the studied samples were ground and exposed to the simulated conditions of the upper gastrointestinal tract, including digestion with porcine enzymes (salivary α-amylase, pepsin, and pancreatin) and removal of the digestion products by dialysis. The remaining non-digestible material was then inserted into a colon model incubated with pooled faecal suspension from healthy human donors in strictly anaerobic conditions, as developed by Aura [286]. Samples were taken at time points 0, 2, 4, 6, 8, and 24 hours during the in vitro colonic fermentation.

LC–MS analysis. The liquid chromatography–mass spectrometry analysis was performed in all studies according to Hanhineva et al. [101]. Briefly, liquid chromatography was performed on a 1290 Infinity Binary UPLC (Agilent Technologies). For the reversed-phase separation in Studies I–III, a Zorbax Eclipse XDB-C18 column (dimensions 2.1 × 100 mm, particle size 1.8 µm) was used as the stationary phase. The mobile phase consisted of water (solution A) and methanol (solution B), both containing 0.1% v/v formic acid. The elution gradient profile was as follows (t [min], %B): (0, 2), (10, 100), (14.5, 100), (14.51, 2), (16.5, 2). For the HILIC separation in Study II and IV, an Acquity UPLC BEH Amide column (dimensions 2.1 × 100 mm, particle size 1.7 µm) was used. Here, the mobile phase consisted of 50% v/v aqueous acetonitrile (solution A) and 90% v/v aqueous acetonitrile (solution B). The elution gradient profile was as follows (t [min], %B): (0, 100), (2.5, 100), (10, 0), (10.1, 100), (14, 100).
Mass spectrometry was performed on an Agilent 6540 Q-TOF with a Jet Stream ESI ion source. The fragmentor voltage used was 100 V and scan range 20–1600 m/z. From every precursor scan cycle (400 milliseconds), 4 most abundant precursor ions were automatically selected for MS/MS fragmentation, excluded after 2 acquired MS/MS spectra, and released again from the exclusion list after 0.25 min. The collision energies used for the MS/MS analysis were (±) 10, 20, and 40 V, for compatibility with the METLIN database. Negative ionization mode, particularly suitable for phenolic acids, was used in Studies I–III; positive ionization mode, required for the detection of e.g. betaines, was used in Studies II–IV.

**Data analysis.** Various software were used for the metabolomics data analysis. The peak picking was performed with Agilent MassHunter Qualitative Analysis v7.0 (Studies I–III) and Agilent Profinder (Study IV). The extracted molecular features were exported to Agilent Mass Profiler Professional v2.2 (Studies I–III) for biostatistical analysis. Heat map representations of the results were prepared with Multiple Experiment Viewer v4.9 (Studies I–III), Guineu [287] (Study III), and R code developed in-house (Study IV). Principal component analysis was performed in SIMCA v14 (Study I and II), Agilent Mass Profiler Professional v2.2 (Study III), and R code developed in-house (Study IV). The statistical tests for the determination of differential metabolites included one-way ANOVA with multiple testing correction using Benjamini–Hochberg false discovery rate (Study I and III), one-way ANOVA with p-value significance threshold at 0.01 (Study II), and both parametric ANOVA and t-test with significance threshold at 0.05 for the semi-targeted analysis of betaines in Study IV.

**Metabolite identification and annotation.** We utilised several information sources and methods to identify and putatively annotate the statistically significant differential metabolites. Currently, our in-house database includes 330 reference standards analysed with RP chromatography (containing both retention time and MS/MS spectra) and 420 reference standards analysed with HILIC. In addition, 56 phytochemical standards have been analysed in our platform as part of a European multiplatform coverage test, organised by COST Action FA 1403 POSITIVe (Interindividual variation in response to consumption of plant food bioactives and determinants involved) [76]. In addition to the in-house database, important sources of reference spectra for putative annotations were publicly available mass spectral databases, METLIN in particular, and previously published literature. In Study I and III, the identification process was performed manually with Agilent MassHunter 7.0 software and comparing the MS/MS spectra with existing databases. In Study II, conducted after these, a semi-automated identification process was used with MS-DIAL [142]: the software provided preliminary annotations based on the same databases and these were then confirmed manually. Furthermore, MS-FINDER [288] was used to putatively annotate metabolites for which no experimental spectra were available by comparing the experimental MS/MS data to *in silico* generated MS/MS spectra.
Figure 4.1. The general metabolomics workflow used in this study. The cereal materials included in the studies were exposed to enzymatic bioprocessing, *in vitro* metabolism in a colonic model system, and *in vivo* metabolism in C57BL/6J mice. The samples were analysed with UPLC–QTOF-MS utilizing RP and/or HILIC columns and negative and/or positive ionisation with automated MS/MS data acquisition. The data analysis includes peak picking and alignment and biostatistical analysis to determine the differential metabolites, which then undergo an identification process.
6 METABOLIC PROFILING OF SOURDOUGH FERMENTED WHEAT AND RYE BREAD² (STUDY II)

6.1 ABSTRACT
Sourdough fermentation by lactic acid bacteria is commonly used in bread baking, affecting several attributes of the final product. We analyzed whole-grain wheat and rye breads and doughs prepared with baker’s yeast or a sourdough starter including Candida milleri, Lactobacillus brevis and Lactobacillus plantarum using non-targeted metabolic profiling utilizing LC–QTOF–MS. The aim was to determine the fermentation-induced changes in metabolites potentially contributing to the health-promoting properties of whole-grain wheat and rye. Overall, we identified 118 compounds with significantly increased levels in sourdough, including branched-chain amino acids (BCAAs) and their metabolites, small peptides with high proportion of BCAAs, microbial metabolites of phenolic acids and several other potentially bioactive compounds. We also identified 69 compounds with significantly decreased levels, including phenolic acid precursors, nucleosides, and nucleobases. Intensive sourdough fermentation had a higher impact on the metabolite profile of whole-grain rye compared to milder whole-grain wheat sourdough fermentation. We hypothesize that the increased amount of BCAAs and potentially bioactive small peptides may contribute to the insulin response of rye bread, and in more general, the overall protective effect against T2DM and CVD.

6.2 INTRODUCTION
Increasing evidence is supporting the protective effect of whole-grain cereal consumption against several noncommunicable diseases, such as type 2 diabetes mellitus, cardiovascular disease and colorectal cancer, as well as overall mortality [13, 14, 28]. This has been attributed to cereal dietary fiber and the array of phytochemicals within the fiber matrix [17, 35], both of which interact with gastrointestinal microbiota and undergo transformations, possibly mediating physiological changes [221]. However, the metabolic pathways leading to these effects are still mostly unknown. Among the phytochemical classes, phenolic acids and alkylresorcinols are abundant in the bran section of whole grains [290, 328].

² Adapted from: Koistinen, V. M., Mattila, O., Katina, K., Poutanen, K., Aura, A. M., & Hanhineva, K. (2018). Metabolic profiling of sourdough fermented wheat and rye bread. Scientific reports, 8(1), 5684. Published under a Creative Commons Attribution 4.0 International License.
These compounds have shown antioxidative, antimicrobial and anticancer effects in vitro.

Bread is one of the most important staple foods consumed worldwide and thus serves as a major source of whole-grain cereals [329]. The intake of bread baked from whole-grain rye (Secale cereale L.) has been shown to cause a lowered postprandial insulin response, known as the rye factor [330, 331], via unknown mechanisms. Common wheat (Triticum aestivum L.) is one of the primary sources of dietary fiber in the United States and several other industrialized countries, whereas rye provides an important dietary fiber source in parts of Northern and Eastern Europe. In the baking of nearly all rye breads and several artisanal wheat breads, such as the San Francisco bread, sourdough fermentation by lactic acid bacteria (LAB) and yeasts is used to improve the texture, sensory properties and shelf life of the bread product [203]. Whereas the emergence of industrial baking caused yeast fermentation to become the dominant practice in bread baking, sourdough has gained interest during the last years, not only from the technical and gastronomical perspective but also because of the increased nutritional value and potential health benefits offered by the ancient biotechnological process [201]. Sourdough fermentation is known to cause transformations of lipids and macromolecules [208] and several phytochemicals, such as phenolic acids, folates and sugar-conjugated bioactive compounds [8]. However, the overall metabolism of small molecules during sourdough fermentation has not been studied with a non-targeted method or any other comprehensive chemical analysis.

The aim of this study was to elucidate the changes in the metabolite profile caused by sourdough fermentation as compared to yeast fermentation and to indirectly compare the differences between sourdough-fermented wholegrain wheat and rye breads. Emphasis was given to bioactive compounds, which potentially contribute to the health benefits of whole-grain products. Similar starters, containing Candida milleri, Lactobacillus brevis and Lactobacillus plantarum, were selected for both types of doughs to represent a typical microbial distribution in sourdough. The fermentation conditions for wheat and rye were chosen to represent a typical sourdough process for each cereal. Liquid chromatography–mass spectrometry (LC–MS) -based non-targeted metabolomics was used to detect and identify the discriminatory compounds for each comparison. The proportion of identified compounds was maximized by combining the use of a standard library, database searches, in silico generated mass spectra, and MS/MS fragment motifs associated with certain molecular moieties.

6.3 MATERIALS AND METHODS

Raw materials. Four different types of flours were used in the baking: endosperm wheat flour (manufacturer’s product code: V500P), wholegrain wheat flour (V1700), endosperm rye flour (R700) and wholegrain rye flour, coarse (R1800KA). All flours were obtained from Fazer Mill & Mixes, Finland. The nutritional composition of the flours is presented in Table 6.1.
Preparation of sourdoughs. The microbial starters, their levels and fermentation conditions used in the preparation of sourdoughs are presented in Table 6.2. The selected yeast (*Candida milleri*) and lactic acid bacteria (*Lactobacillus brevis* and *Lactobacillus plantarum*) are commonly found in both rye and wheat sourdoughs [332]. For the rye sourdough, a high proportion of starter and intensive fermentation conditions were selected to produce a high level of acidity, which is desirable for obtaining a good crumb structure in rye bread [333]. For the wheat sourdough, a lower level of starter and milder fermentation conditions were selected to produce a moderate acidity level, which has been related to improved wheat sourdough bread flavour and volume [334]. The microbial strains were obtained from the culture collection of VTT. The starters were prepared by refreshing the microbial strains twice in succession in general edible medium [335] for 24 h at 30 °C. The cells were collected from the cultures by centrifugation and suspended in sterile water to obtain desired cell concentrations. Sourdoughs were prepared by mixing 1 kg of flour with 1.5 l of water and starter suspension by hand and incubating the mixture in a covered container. After the fermentation, a sample was taken from the sourdough for the analysis of microbial growth and a sample was frozen for a subsequent analysis of acidity. The viable counts of lactic acid bacteria were determined using plate count technique on MRS agar (Oxoid, Basingstoke, UK). Yeasts were enumerated on YM agar (Difco laboratories, Detroit, USA). Acidity of the sourdoughs (pH and total titratable acidity, TTA) was analysed according to a standard method [336]. The fresh sourdoughs were immediately used for bread baking.

Table 6.1. The content (as is) of protein, dietary fibre and fat in the flours as provided by the manufacturer.

<table>
<thead>
<tr>
<th></th>
<th>Endosperm wheat</th>
<th>Wholegrain wheat</th>
<th>Endosperm rye</th>
<th>Wholegrain rye</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein</td>
<td>12</td>
<td>12</td>
<td>5.5</td>
<td>9.1</td>
</tr>
<tr>
<td>Dietary fibre</td>
<td>4.4</td>
<td>13</td>
<td>9.6</td>
<td>18</td>
</tr>
<tr>
<td>Fat</td>
<td>1.7</td>
<td>3.2</td>
<td>2.3</td>
<td>2.3</td>
</tr>
</tbody>
</table>

Table 6.2. The microbial starters and their dosages (cfu/g sourdough in the beginning of the fermentation), the fermentation conditions, the acidity test results and the microbial count results (cfu/g) of the sourdoughs at the end of the fermentation.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Wholegrain wheat</th>
<th>Wholegrain rye</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Candida milleri</em> C-96250</td>
<td>$10^6$</td>
<td>$10^7$</td>
</tr>
<tr>
<td><em>Lactobacillus brevis</em> E-95612</td>
<td>$10^7$</td>
<td>$10^9$</td>
</tr>
<tr>
<td><em>Lactobacillus plantarum</em> E-78076</td>
<td>$10^7$</td>
<td>$10^9$</td>
</tr>
<tr>
<td>Fermentation time and temperature</td>
<td>12 h, 24 °C</td>
<td>20 h, 32 °C</td>
</tr>
<tr>
<td>pH</td>
<td>4.7</td>
<td>3.8</td>
</tr>
<tr>
<td>Total titratable acidity (TTA), ml</td>
<td>8</td>
<td>15</td>
</tr>
<tr>
<td>LAB count (end of fermentation)</td>
<td>$10^9$</td>
<td>$10^9$</td>
</tr>
<tr>
<td>Yeast count (end of fermentation)</td>
<td>$10^7$</td>
<td>$10^8$</td>
</tr>
</tbody>
</table>
Preparation of breads. Six different breads were prepared by straight dough or by sourdough baking process at VTT Technical Research Centre of Finland Ltd., Espoo (Table 6.3). For each bread, two replicates were baked on separate days to account for possible changes in the fermentation process. The water content of the dough for each of the flours was determined by a farinograph and/or adjusted by test baking. In the sourdough breads, 33% of the total flour was sourdough fermented. The water content of the dough was the same for the sourdough breads and the corresponding straight dough breads. The wheat breads were moulded mechanically (by a conical rounder and a long moulder) and the rye breads by hand (due to dough stickiness). After moulding, the dough pieces were placed in fat-sprayed aluminium pans for proofing. One of the pans was left without spray fat. After proofing, a dough sample was taken from the pan without spray fat in order to avoid extra fat in the proven dough sample. After baking, the breads were left to cool down for 2–4 h before an analysis of volume and texture, except for the wholegrain rye breads, which were analysed one day after the baking because the breads were too sticky to be analysed on the baking day. A slice was cut from the middle of each bread, cut further into cubes of ca. 1 cm in size, and stored frozen in sealed plastic bags until the sample preparation for the analysis.

Table 6.3. The recipes (as percentage of flour weight) and the baking process of the breads.

<table>
<thead>
<tr>
<th></th>
<th>Straight dough breads</th>
<th>Sourdough breads</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flour</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>Sugar</td>
<td>2.0</td>
<td>2.0</td>
</tr>
<tr>
<td>Sourdough</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>of which flour</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>of which water</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Water</td>
<td>61.0</td>
<td>70.0</td>
</tr>
<tr>
<td>Yeast</td>
<td>3.0</td>
<td>3.0</td>
</tr>
<tr>
<td>Salt</td>
<td>1.2</td>
<td>1.2</td>
</tr>
<tr>
<td>Mixing</td>
<td>2 min slow + 4 min fast</td>
<td></td>
</tr>
<tr>
<td>Resting</td>
<td>20 min 28 °C / 80% RH</td>
<td>60 min 28 °C / 80% RH</td>
</tr>
<tr>
<td>Dough piece weight</td>
<td>400 g</td>
<td>500 g</td>
</tr>
<tr>
<td>Moulding</td>
<td>mechanically</td>
<td>by hand</td>
</tr>
<tr>
<td>Resting</td>
<td>8 min</td>
<td>–</td>
</tr>
<tr>
<td>Proofing</td>
<td>50 min 37 °C / 8% RH</td>
<td>60 min 37 °C / 80% RH</td>
</tr>
<tr>
<td>Steaming in oven</td>
<td>15 s</td>
<td></td>
</tr>
<tr>
<td>Baking</td>
<td>225 °C 20 min</td>
<td>240 °C 10 min + 220 °C 35 min</td>
</tr>
</tbody>
</table>
Sample preparation. Samples were obtained from each type of bread ($n = 6$), dough ($n = 6$) and flour ($n = 4$). The frozen samples (−80 °C) were cryoground into fine powder using a tissue homogenizer (TissueLyser II, Qiagen, Hilden, Germany). The grinding frequency was set at 20 s$^{-1}$ and the duration at 40 s. The ground sample was inspected visually, and the procedure was repeated in the case of any unground pieces. Approximately 100 mg of each sample was weighed into Eppendorf tubes; at this stage, four technical replicates were taken from the flour samples and three replicates from the dough and bread samples into separate tubes. The metabolite extraction solvent was produced by mixing HPLC gradient grade methanol, MilliQ water and 80% formic acid using a MeOH:H₂O:HCOOH ratio of 80:19.9:0.1 v/v/v. The solution had a pH of 3.6. The extraction solvent was added into the tubes in a ratio of 300 µl to 100 mg of sample. The tubes were vortexed, sonicated for 15 min, vortexed again, and centrifuged for 10 min at 13,000 rpm. The supernatant was collected and filtered (Acrodisc CR 13 mm syringe filter with 0.2 µm PTFE membrane). The filtrate was then transferred to HPLC sample vials.

LC–MS/MS analysis. The liquid chromatography–mass spectrometry was performed on a 1290 Infinity Binary UPLC coupled with a 6540 UHD Accurate-Mass Q-TOF (Agilent Technologies, Santa Clara, CA, USA) as described previously by Hanhineva et al. [84] Briefly, a Zorbax Eclipse XDB-C18 column was used for the reversed-phase separation and an Aquity UPLC BEH amide column (Waters, Milford, MA, USA) for the HILIC separation. After each chromatographic separation, the ionization was carried out using jet stream electrospray ionization (ESI) in the positive and negative mode, yielding four data files per sample. The collision energies for the MS/MS analysis were chosen as 10, 20 and 40 V, for compatibility with the spectral databases.

Statistical analysis and compound identification. The molecular features were extracted from the data by using MassHunter Qualitative Analysis version 7.0 (Agilent Technologies) and the biostatistical analysis was performed in Mass Profiler Professional version 2.2 (Agilent Technologies), as described previously [74]. The raw values of the peak areas were used in determining the relative levels of the compounds in different samples. The $p$ values were calculated using one-way ANOVA and the fold changes as the ratio of the average raw peak areas of sourdough fermented samples to yeast fermented samples, using a negative inverse value in case of a ratio below 1 (level decrease). After the statistical analysis, filtering was performed in Microsoft Excel to extract the list of statistically significant peaks for each comparison. Fold change was set at ≥2 (overexpression in sourdough samples), $p$ value at <0.01, RSD at ≤30% (variation within sample group), and average signal intensity (in the sample group being compared) at ≥200 000. Principal component analysis (PCA) was performed in SIMCA version 14 (Umetrics AB) using Pareto scaling. Hierarchical Clustering method in Multiple Array Viewer version 4.9.0 (TM4 Software Suite) was used to create the heat map of the significant identified compounds.

The accurate masses, retention times, and MS/MS fragmentation patterns of the detected statistically significant compounds were used in comparison with an in-
house standard library, previously published literature, and freely available MS spectral databases. The tentative annotation of compounds based on spectral database searches was performed in MS-DIAL [142]. For some of the compounds with no spectral database matches, tentative IDs were given based on in silico generated MS/MS spectra acquired from molecular database entries using MS-FINDER software [288]. For the identification of small peptides, fragment motifs specific for amino acid residues were determined from the identifications based on MS/MS data and used in the further identification process. The fragment motifs observed in at least two peptides and used in the identifications are presented in Table 6.4.

Table 6.4. Proposed MS/MS fragment motifs and their tentative structures associated with amino acid residues of small peptides in mass spectra from LC–MS with electrospray ionization in the positive mode (ESI+).

<table>
<thead>
<tr>
<th>Fragment m/z</th>
<th>Amino acid</th>
<th>Tentative fragment structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>44.050</td>
<td>(unspecific)</td>
<td></td>
</tr>
<tr>
<td>55.053</td>
<td>valine</td>
<td>unknown</td>
</tr>
<tr>
<td>56.049</td>
<td>(unspecific)</td>
<td></td>
</tr>
<tr>
<td>60.044</td>
<td>serine</td>
<td>side chain</td>
</tr>
<tr>
<td>60.055</td>
<td>arginine</td>
<td>guanidine group</td>
</tr>
<tr>
<td>69.071</td>
<td>leucine, isoleucine</td>
<td>branched hydrocarbon chain</td>
</tr>
<tr>
<td>70.065</td>
<td>(unspecific)</td>
<td></td>
</tr>
<tr>
<td>72.080</td>
<td>valine</td>
<td>branched hydrocarbon chain</td>
</tr>
<tr>
<td>74.057</td>
<td>threonine</td>
<td>side chain</td>
</tr>
<tr>
<td>84.080</td>
<td>lysine, leucine, isoleucine</td>
<td>side chain</td>
</tr>
<tr>
<td>86.095</td>
<td>leucine, isoleucine</td>
<td>branched hydrocarbon chain</td>
</tr>
<tr>
<td>101.072</td>
<td>threonine</td>
<td>side chain</td>
</tr>
<tr>
<td>110.071</td>
<td>histidine</td>
<td>side chain</td>
</tr>
<tr>
<td>116.070</td>
<td>proline, arginine</td>
<td>molecular ion (proline)</td>
</tr>
<tr>
<td>117.032</td>
<td>valine</td>
<td>molecular ion</td>
</tr>
<tr>
<td>120.078</td>
<td>phenylalanine</td>
<td>side chain</td>
</tr>
<tr>
<td>130.048</td>
<td>glutamine</td>
<td>side chain</td>
</tr>
<tr>
<td>136.078</td>
<td>tyrosine</td>
<td>side chain</td>
</tr>
<tr>
<td>147.077</td>
<td>glutamine</td>
<td>molecular ion</td>
</tr>
<tr>
<td>175.120</td>
<td>arginine</td>
<td>molecular ion</td>
</tr>
</tbody>
</table>

6.4 RESULTS

Effect of sourdough on the metabolic profiles. Principal component analysis (PCA) of the raw data, including the two first components and all the detected molecular features, showed a clear separation of the metabolic profiles of the flours from the rest of the sample groups (doughs and breads) along the first component of the PCA (Figure 6.1). It also revealed a pronounced separation of the sourdough fermented whole-grain rye bread (WWSB) and dough (WWSD) from the other sample groups along the second component. In contrast, yeast fermented whole-grain rye bread and
dough showed no major difference to similarly prepared bread and dough produced from sifted rye flour (with bran section removed). Sourdough fermented whole-grain wheat bread and dough, however, did not show as large changes in their metabolite profiles as rye when compared with their yeast fermented counterparts. In wheat, the difference between the overall metabolite profiles depended mostly on the whole-grain content of the samples. The changes in the profiles occur mostly during the baking of the dough, as indicated by the proximity of the dough and bread samples of similar bread types in the PCA.

Figure 6.1. Principal component analysis (PCA) of the bread samples. This figure contains the first two principal components and their scores $t_1$ and $t_2$, which explain 21% and 16% of the variation within the data, respectively. WF = white wheat flour, WWF = whole-grain wheat flour, RF = refined rye flour, WRF = whole-grain rye flour, WD = white wheat dough, WB = white wheat bread, WWD = whole-grain wheat dough, WWB = whole-grain wheat bread, WWSD = whole-grain wheat sourdough, WWSB = whole-grain wheat sourdough bread, RD = refined rye dough, RB = refined rye bread, WRD = whole-grain rye dough, WRB = whole-grain rye bread, WRSD = whole-grain rye sourdough, WRSB = whole-grain rye sourdough bread, QC = quality control from pooled samples.
Metabolites with increased levels. When comparing data from sourdough and yeast fermented whole-grain rye breads, 711 molecular features fulfilled the criteria of significant increase in sourdough rye bread ($p < 0.01$, fold change (FC) $\geq 2$). Correspondingly, 212 features fulfilled the same criteria in the sourdough fermented wheat bread. Out of these, 79 features had significantly increased levels in both cereal type sourdough breads (Figure 6.2). After the identification process and removal of redundant ions from the data from different ionization modes and columns, 118 distinct compounds were identified with a standard or putatively identified among the significant compounds, out of which 110 were significant in the rye bread comparison (sourdough vs. yeast fermented), 49 in the wheat bread comparison, and 41 in both comparisons. Table S1 (available from https://doi.org/10.1038/s41598-018-24149-w) shows the identified compounds along with their identification and statistical data. The identified compounds and their relative levels across all studied samples were visualized in a heat map (Figure 6.3).

Figure 6.2. The number of molecular features and identified compounds with significantly increased or decreased levels in the whole-grain sourdough rye and wheat breads compared to their yeast-fermented counterparts.

Figure 6.3 (next page). Heat map of the identified compounds with significant level increase (orange) or decrease (blue) ($p < 0.01$, FC $\geq 2$) in sourdough wheat and/or rye bread. Data is included from all the studied samples.
Overall, the identified differential compounds of sourdough fermentation fell into several categories of potentially bioactive chemicals. In sourdough rye, the most notable group of identified compounds was amino acids and their derivatives, including metabolites of amino acids and di-, tri- and tetrapeptides (Table S1). The amino acids included eight proteinogenic amino acids (asparagine, glutamic acid, isoleucine, leucine, lysine, phenylalanine, methionine, and tryptophan) and four other amino acids (citrulline, homocitrulline, ornithine, and saccharopine). Notably, two branched-chain amino acids (BCAA), leucine and isoleucine, as well as leucic acid (2-hydroxyisocaproic acid), a metabolite of leucine, were among the differential compounds. The increase in the levels of leucine and isoleucine was 17-fold and 10-fold, respectively. The increased metabolites of amino acids also included 2-hydroxyisovaleric acid (from BCAAs), tryptophan metabolites indole-3-lactic acid and 3-phenyllactic acid, and tyrosine metabolites 4-hydroxyphenyllactic acid and tyramine. Out of the 70 small peptides putatively identified, 64 fulfilled the significance criteria in rye and 62 contained one or several BCAA residues. As indicated by the heat map (Figure 6.3), most of the small peptides and several amino acids and their metabolites were highly specific to whole-grain rye sourdough and the corresponding bread. Three microbial metabolites of phenolic acids (dihydroferulic acid, dihydrocaffeic acid and dihydroxysinapic acid), phenolic acid derivatives (feruloylagmatine and p-coumaroylputrescine), six sugars or sugar derivatives, two phosphatidylcholines [PC(18:2/18:2) and PC(18:2/18:3)], and two fatty acids were among the increased compounds. Other identified potentially bioactive compounds with significantly increased levels were 2-benzoxazolinone (benzoxazinoid), 2-hydroxyvaleric acid, isorhamnetin (flavonoid), N-acetylspermidine, and phenylethanolamine. The identified and unidentified compounds were plotted based on their retention time, fold change and average signal intensity to highlight the most important differential compounds found in the analysis (Figure 6.4).
In wheat sourdough, two amino acids (citrulline and saccharopine), 29 small peptides and six amino acid metabolites (the same as in rye sourdough) had significantly increased levels (\( p < < 0.01, \text{FC} \geq 2 \) (Figure 6.4B). In contrast to rye, none of the proteinogenic amino acids were among the differential compounds. One small peptide, YQK, was significantly increased only in wheat sourdough. Phenolic acids and their microbial metabolites behaved similarly in wheat sourdough compared to rye (Figure 6.3). No phosphatidylcholines were significantly changed in wheat.

**Decreased and unchanged metabolites.** We also investigated the compounds that had significantly decreased levels in the sourdough bread samples. In the rye breads, we found 352 molecular features fulfilling the same inclusion criteria as for the features with increased levels, including 69 identified metabolites (Table S1). 143 features (34 identified compounds) fulfilled the criteria in the wheat bread samples and 49 features (18 identified compounds) in both wheat and rye samples. The identified decreased metabolites in either rye or wheat sourdough included phenolic acids (caffeic, ferulic, \( p \)-coumaric and sinapic acid), five nucleosides involved in nucleic acid synthesis (adenosine, cytidine, guanosine, thymidine and uridine), nucleoside derivatives and nucleobases (adenine, cytosine, guanine, and thymine), amino acid–derived betaines (trimethyllysine and valine betaine), \textit{myo}-inositol, two phosphatidylcholines \([\text{LysoPC}(15:0)\) and \(\text{PC}(16:0/18:3)\)], mono- and polysaccharides, and oxidized fatty acids (shown for rye in Figure 6.4C). There were differences between the wheat and rye sourdoughs: the two amino acid-derived betaines and most of the fatty acids, nucleobases and sugars fulfilled the inclusion criteria only in rye. As shown in Figure 6.3, the levels of nucleosides and their derivatives and nucleobases
were increased by yeast fermentation while the relative levels were low in both rye and wheat flours and sourdoughs.

We identified several compounds belonging to the major phytochemical classes known to exist in whole-grain cereals that were not significantly ($p < 0.01$) increased or decreased in their levels based on the chosen thresholds in sourdough fermented breads. Nine alkylresorcinols (17:0, 17:1, 19:1, 19:2, 21:1, and the levels of four oxidated alkylresorcinols) were significantly ($p < 0.05$) increased in the sourdough rye bread (between 1.15 and 3.49; Table S2). None of the levels of alkylresorcinols identified from the data were significantly changed in wheat sourdough. In total, twelve amino acid-derived betaines were identified from the data, many of them previously unreported from cereals (4-aminobutyric acid betaine, glutamic acid betaine, glutamine betaine, histidine betaine, isoleucine betaine, pipecolic acid betaine, and trimethyllysine). In addition to trimethyllysine and valine betaine, the levels of histidine betaine and trigonelline were significantly decreased in rye sourdough (FC −1.89 and −1.49, respectively) and isoleucine betaine increased in both rye and wheat sourdough (FC 3.57 and 2.97, respectively). The levels of detected lignans were generally too low for reliable identification; the only putatively identified lignan, buddlenol C, showed significant minor decrease in levels ($p < 0.001$, FC −1.27) in rye sourdough and significant minor increase in levels ($p < 0.05$, FC 1.15) in wheat sourdough.

6.5 DISCUSSION

The principal component analysis indicated that the most substantial difference in the metabolic profiles of the analyzed samples was between the flours and all the other samples (including the doughs and breads). Clear differences were seen between wheat and rye, between whole-grain and processed samples, and between rye sourdoughs and straight doughs, highlighting the significance of food processing in the final biochemical composition of the products. The effect of sourdough alone was considerable, especially in rye, where more than 700 molecular features had increased levels with the chosen inclusion criteria and over 350 decreased levels with high significance. These changes are likely contributing to the differences in the sensory properties and the potential metabolic effects of sourdough bread. Many of the observed changes in the metabolite levels were specific to rye sourdough; overall, sourdough fermentation of whole-grain wheat produced considerably less significantly changed metabolites and a less separated metabolic profile from yeast-fermented wheat dough and bread in the PCA. The main reason behind this is likely the different conditions chosen for the two sourdoughs, selected to represent widely consumed breads, with the higher temperature and longer fermentation time of rye sourdough allowing a more extensive metabolism. In contrast to rye, only mild acidity is accepted by consumers in wheat sourdough bread, thus requiring differences in the fermentation process [334]. The appearance of certain metabolites exclusively in rye sourdough may also be explained by the wider range of
compounds in rye available for microbial metabolism, differences between wheat and rye flour matrices, the development of LAB ecology during the baking, and different endogenous enzymatic activity in the two species.

Sourdough fermentation considerably increased the levels of branched-chain amino acids leucine and isoleucine, BCAA metabolites as well as several small peptides containing BCAAs. This effect was more prominent in rye than wheat sourdough most likely due to intensive proteolysis in the acidic rye sourdough utilized in this study. As has been shown previously, the consumption of rye bread decreases the postprandial insulin response without a decrease in glucose response due to an unknown mechanism independent from the dietary fiber content of the bread [50, 330]. BCAAs are known to activate the mTORC1 signaling pathway in skeletal muscle in a similar manner to exercise, and this in turn leads to uncoupling of the insulin signaling [337]. This may contribute to the decreased insulin response seen in vivo. However, the role of BCAAs in insulin metabolism is more complex than this, as their increased circulating levels are associated with obesity-related insulin resistance, possibly because of an overload of BCAA catabolism [338]. In a study by Moazzami et al., higher fasting concentrations of leucine and isoleucine were correlating with a higher insulin response after the intake of all types of study breads [339]. In contrast, phenylalanine and methionine, which were also among the significantly increased amino acids in sourdough rye in the current study, were the main metabolites associated with a lowered insulin response after 60 minutes of sourdough rye bread intake [340].

Several small peptides have been reported to possess potential antioxidant and antihypertensive activity [342, 343]. These characteristics seem to be governed by the presence of certain amino acids in the peptide sequence; e.g. leucine may increase both the radical scavenging activity and ACE inhibition of the peptide [344]. Specifically, 28 peptides that had increased levels in sourdough (marked with an asterisk in
Table S1) have been included in the database of antihypertensive peptides [345] and VKL as antioxidant [342].

The main phenolic acids present in rye and wheat – ferulic, caffeic, p-coumaric, and sinapic acid – had significantly lower levels in both rye and wheat sourdough. Correspondingly, the levels of several known microbial metabolites of phenolic acids [177, 247] were increased after sourdough fermentation, indicating that they were likely metabolized by the LAB strains. This observation is in agreement with in vitro studies, where various LAB strains were shown to metabolize all the main phenolic acids [346, 347]. These microbial metabolites have different absorption and metabolic characteristics than their precursors [348], which may have significance regarding the bioactivity of these compounds. Although rye and wheat are not abundant sources of flavonoids, we observed an increase in the levels of isorhamnetin and a corresponding decrease in the levels of isorhamnetin-3-O-hexoside, indicating the release of the flavonoid aglycone from its glycoside by the bacteria. The levels of most alkylresorcinols increased in rye sourdough. Since the increase was absent in wheat sourdough, the observation is most likely not originating from analytical factors, such as the relatively poor solubility of alkylresorcinols in commonly used extraction solvents. This result is in disagreement with previous studies, where alkylresorcinols have been reported to either decrease [74, 209] or exhibit only minor changes [205] after sourdough fermentation. The increase of alkylresorcinol levels seen here in only sourdough rye might be explained by different capabilities of LAB strains to release or metabolize these compounds and the softer matrix in rye bran compared to wheat [74].

Interestingly, nucleobases, nucleosides and their derivatives had lower levels in sourdough wheat and rye compared to yeast fermented samples. Since their relative levels were found to be low in flours as well, they are likely metabolites of yeast fermentation. Little is known about the biological significance of these compounds regarding dietary intake; among the few reported effects is that they may act as immunomodulators in infants receiving nucleosides in breast milk [349]. Four species of phosphatidylcholines had significantly changed levels in rye sourdough. The compounds with observed increase in their levels contain polyunsaturated fatty acids (PUFAs), likely linoleic acid and alpha- or gammalinolenic acid. The dietary intake of ω-3 phosphatidylcholines has been shown to improve fatty acid and glucose metabolism in rats [350], but on the other hand, the colonic microbial metabolism of the choline group in phosphatidylcholines may result in adverse cardiovascular effects [351]. We could not determine in the current study whether the identified phosphatidylcholine species contain ω-3 or ω-6 PUFAs; in addition, more research is needed to determine the association of polyunsaturated phosphatidylcholine intake with the risk of CVD and diabetes.

While non-targeted metabolomics is ideal for a wide-scale investigation of the metabolic profile of any biological sample, it is limited by the incomplete availability of spectral references for the reliable annotation of all statistically significant compounds, and this holds true especially for metabolite-rich sample matrices, such as
the ones studied in this work. The high relative number of unknowns may cause bias towards more well-known types of compounds, for which more reference data is available. Therefore, future targeted studies are warranted to further investigate the changes occurring in these metabolites as well as providing confirmation to the identifications. There is some uncertainty in annotating small peptides due to limitations in separating the signals originating from leucine and isoleucine and deducing the correct order of the amino acid residues in the peptide. Peptides larger than three amino acid residues have limited reference spectra available. The different fermentation conditions in wheat and rye sourdoughs inevitably affect the metabolic profiles observed in the current study; however, the differences in the profiles are relevant regarding the potential health implications, since the studied breads represent similar types of breads as normally consumed. The selection of LAB strains may affect the metabolic profiles or sourdough fermentation depending on the array of enzymes present in each strain. Here, we aimed to use one representative sourdough starter; studying the effect of the selection of strains on the same metabolic profiles is warranted for the future.

The non-targeted metabolomics approach provided wide insight into the metabolic profile of sourdough fermentation. We hypothesize that sourdough fermentation contributes to the beneficial health effects of whole grains by increasing the amount of several bioactive compounds, such as BCAAs, small peptides, and microbial phenolic acid metabolites, in the baked products. The effect is more profound in rye sourdough compared to wheat, likely attributable to the more extensive metabolism occurring in a typical rye sourdough and suggesting a potential contributor to the rye factor. The current study indicates the potential for future research to reveal the molecular basis of physiological signals caused by wholegrain bread intake, and in more general, of the observed protective effect of whole grains against non-communicable diseases.

6.6 ACKNOWLEDGEMENTS

We thank Dr Hiroshi Tsugawa from RIKEN Yokohama Branch for assistance in the use of MS-DIAL and MS-FINDER software, Dr Riikka Juvonen from VTT for instructing the preparation of the sourdough starters, Miia Reponen from University of Eastern Finland for assistance in the LC–MS analysis, and MSc Marco Russo from University of Naples Federico II for compiling the in-house database into format supported by MS-DIAL. This work was supported financially by Academy of Finland, Lantmännen Research Foundation, and Biocenter Finland.

Data availability. The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

Supplementary material. Table S1 and S2 are available as electronic supplementary material at https://doi.org/10.1038/s41598-018-24149-w.
9 SYNOPSIS OF THE RESULTS

The aim of Study I was to investigate the effect of enzymatic bioprocessing on the phytochemicals of rye bran, which was added into white wheat bread. The comparison was made against white wheat bread, white wheat bread supplemented with unprocessed rye bran, and sourdough-fermented whole-grain rye bread. As revealed by nontargeted LC–MS metabolite profiling, the bioprocessing induced changes in the phytochemical profile dissimilar to those caused by sourdough fermentation: expectedly, the levels of free phenolic acids were significantly increased, up to 108-fold in ferulic acid, but also the levels of several benzoxazinoid and flavonoid aglycones and monoglycosides increased, suggesting the release of sugar units from these compounds. Some compounds were characterised for the first time from rye bran, including DIM2BOA, 3,4,5-trimethoxycinnamic acid, and ethyl p-coumarate.

Study II aimed to further elucidate the overall metabolite profile of sourdough fermented whole-grain rye and wheat bread compared to their yeast-fermented counterparts, analysed with the nontargeted LC–MS method. Sourdough extensively altered the metabolic fingerprint of rye, causing an increase in the levels of microbial metabolites of phenolic acids, branched-chain amino acids (BCAA), amino acid metabolites, and BCAA-containing small peptides, while the levels of phenolic acid precursors decreased. In wheat, typically prepared with less intensive sourdough fermentation compared to rye, the effect of sourdough was less prominent. In total, 118 metabolites with increased levels (\(p < 0.01, \text{fold change} \geq 2\)) and 69 metabolites with decreased levels were identified or putatively annotated from rye and wheat sourdough.

In Study III, the purpose was to study the effect of bran bioprocessing on the gut microbial metabolism of phenolic acids using an in vitro model of the human gastrointestinal system and analysing the samples with two complementary methods, GC × GC–TOF-MS and UPLC–QTOF–MS. In contrast to previously published results from wheat, enzymatic bioprocessing did not significantly influence the appearance of microbial metabolites, despite that the phenolic acid precursors were more extensively released from the fibre matrix. The microbial metabolic pathways of diferulic acids were reviewed and additional metabolites were suggested to be linked to the metabolism based on the results, including \(p\)-coumaric acid, vanillic acid, 3-(4-hydroxyphenyl)propionic acid, 4-hydroxyphenylacetic acid, and 4-methylcatechol.

The objective of Study IV was to determine the effect of bran-enriched diet on the gut microbial composition of C57BL/6J mice and to discover whether gut microbiota is responsible for the metabolism of dietary glycine betaine by additionally studying conventional and germ-free C57BL/6NTac mice, in vitro model of the human gastrointestinal system incubated with rye bran, and the whole-grain breads from Study II. The levels of betainised compounds increased in the colon contents of the C57BL/6J
mice fed with bran-enriched diets. The diets also induced changes in the gut microbiota of the mice, analysed with 16S rRNA gene sequencing, increasing the proportion of several bacterial taxa associated with beneficial health effects, including Akkermansia, Bifidobacterium, Coriobacteriaceae, Lactobacillus, Parasutterella, and Ruminococcus. Conventional C57BL/6NTac mice had significantly higher levels of betainised compounds in the gut tissues compared to germ-free mice, although the compounds were not completely missing from the germ-free mice. High levels of glycine betaine were observed in the studied whole-grain cereal samples in contrast to other betainised compounds, which were present in low levels or trace amounts in the samples.
10 GENERAL DISCUSSION

This thesis work was part of a research project which was aiming to identify, characterise and validate biomarkers related to food processing, digestion, absorption, and metabolism and associated with the metabolic effects of whole grain intake. The studies included in this thesis were focusing on a) elucidating the phytochemical composition of the whole-grain products of interest (mainly wheat and rye breads with various processing) in Study I and II and b) the microbial metabolism of phytochemicals in an in vitro gastrointestinal model (Study III and IV) and in mouse models (Study IV). The comprehensive discussion of these four studies are included in Chapters 5–8; this discussion focuses on the utilisation of the results and methodological aspects from a critical perspective.

10.1 UTILISATION AND RELEVANCE OF THE RESULTS

Study I showed that enzymatic bioprocessing significantly increased the bioaccessibility (amount of freely available compounds) of phenolic acids contained in rye bran, which was added to white wheat bread. Unexpectedly, also the levels of several benzoxazinoids and flavonoids increased markedly. Bran bioprocessing thus seems to offer an interesting possibility in food technology to increase the absorption of phytochemicals already in the small intestine, since they do not require release from the fibre matrix by colonic microbiota. However, the physiological mechanisms of phenolic acids are not well known, and it remains unsettled whether the microbial metabolites of phenolic acids are more bioactive than the precursors present in cereals, which would make the colonic metabolism an essential part of mediating the potential health benefits of phenolic acids. In other words, it cannot be confirmed by this study whether bran bioprocessing is a beneficial food processing technique from the health perspective or whether the phenolic acids in whole grains should be left in the fibre matrix to be exposed to microbial metabolism and support the growth of beneficial gut microbes. Furthermore, the production of enzymes should be economically feasible to produce functional foods containing the bioprocessed bran with sufficient market potential. In Study I and II, sourdough fermentation also caused a major shift in the metabolite profile of rye and wheat, increasing the levels of microbial metabolites of phenolic acids, other phytochemicals, branched-chain amino acids (BCAAs), and small peptides with a high proportion of BCAAs and potential biological activity. The results suggest that sourdough fermentation is a contributing factor to the health benefits of whole grains, providing one possible explanation to the observed rye factor (Chapter 2.1.2). Since sourdough also has beneficial effects on the sensory and technological properties, its use should be promoted in the production of breads and other cereal products.
In study III, no major difference between the in vitro colonic microbial metabolism of bioprocessed and unprocessed rye bran was observed, contrary to results from wheat bran by Mateo Anson et al. [216], using a similar model. This may indicate that the rye bran matrix, which is dissimilar to wheat, behaves differently after bioprocessing and during colonic fermentation, allowing a sufficient amount of phenolic acids to access the colonic metabolism regardless of the enzymatic bioprocessing, which releases some of the compounds to be readily absorbed in the small intestine. Thus, enzymatic bioprocessing of rye bran may both increase the bioavailability of phenolic acid precursors, such as ferulic and p-coumaric acid, while still leaving a proportion of these compounds unreleased to be transformed into microbial metabolites with their own potential health benefits. Enzymatic bioprocessing may therefore have more applications in rye bran fortification, which can easily increase the nutritional value of processed foods, such as white wheat bread.

The results from Study IV suggest that betainised compounds are produced most likely from glycine betaine by gut microbiota in mice and confirm previous findings on the influence of whole-grain cereals on the gut microbial composition. Among the compounds was 5-aminovaleric acid betaine (5-AVAB), which we have previously shown to affect the energy metabolism of mouse cardiomyocytes [113]. Betaines thus seem to be a novel group of bioactive compounds that may mediate the beneficial effects of the diet–microbiota interaction. The results still require verification by mechanistic studies including isotope labelled glycine betaine for the determination of the metabolic pathways and incubation with selected strains of bacteria to elucidate the specificity of the metabolism.

The reported LC–MS characteristics of all the identified and putatively annotated metabolites in Study I (n = 60), Study II (n = 187), Study III (n = 30), and Study IV (n = 17) provide additional information for the future characterisation of whole grain phytochemicals and other metabolites. The studies also confirm the applicability of nontargeted LC–MS metabolomics in the analysis of phytochemicals from whole grains, as discussed previously by Pekkinen in her dissertation [416]. Although the inclusion of only wheat and rye in the studies can be justified by them being the main cereal species used in bread production, widening the scope of research to other cereals and pseudocereals would have undoubtedly offered additional interesting observations and a wider insight into the inter-species differences in the changes caused by the food processing techniques and metabolism studied herein. However, this may have not been feasible within the resource and time frame reserved for the studies.

10.2 STRENGTHS AND LIMITATIONS OF THE METHODOLOGY

Nontargeted metabolomics. Unlike in targeted metabolite analyses and metabolite profiling, where a previously known or otherwise limited set of metabolites is often quantitatively measured, nontargeted metabolomics is a hypothesis-generating
method, where differential metabolites between e.g. treatment and control are detected in a semi-quantitative manner, determined by statistical analysis, and finally identified. The differential metabolites are unknown in the beginning of the analysis and may contain novel compounds, which gives potential for new important findings and directions for further research. The downside of this approach is that unless combined with additional targeted mechanistic studies, the interpretation of the results is limited to associations between the metabolites and the outcomes and does not yet confirm the metabolic pathways or other causalities responsible for the outcomes. This limitation belongs to the nature of untargeted metabolomics and does not diminish the need for untargeted studies, since novel findings are often made by chance.

Another limitation in the methodology is the analytical coverage of the LC–MS method. While the use of both RP and HILIC columns and positive and negative ionisation expand the coverage of phytochemical classes, some of the them may be underrepresented in the results due to other factors, such as the uniform sample preparation step using aqueous methanol. Out of the groups of phytochemicals presented in Chapter 2.2, carotenoids, lignans, phytosterols and tocols were nearly absent among the identified compounds, suggesting that they were not properly extracted from the samples by methanol or they were beyond the analytical range of the LC–MS method. In the interlaboratory coverage test, our method was able to detect and identify the carotenoid, lignan, and tocol standards dissolved in appropriate solvents [76], suggesting that extraction is the main limiting factor in the analytical coverage. To increase the coverage, a two-phase extraction method with a non-polar organic solvent would allow more lipophilic metabolites to be extracted and analysed [259]; however, it would double the number of analysed samples (and produced data files), thus creating another bottleneck in terms of instrument time and the effort needed for data analysis. Arguably, to be called a true metabolomics approach, the analytical method used should not be severely limited in its coverage of chemical classes, which may be the case in GC–MS limited to small and relatively volatile compounds and LC–MS methods optimized for certain compounds of interest.

Statistical testing is challenging in nontargeted metabolomics due to the vast number of metabolites compared to the number of samples in a typical dataset [267, 271]. The number of false positive discoveries can be reduced with multiple testing correction, as performed in Study I and II using Benjamini–Hochberg false discovery rate; however, in Study III, another approach was chosen by lowering the significance threshold of the raw $p$ value to 0.01. This means that out of the 118 metabolites found as potential biomarkers of sourdough fermentation in Study III, one or a few metabolites appearing as significant may be expected to be false positive discoveries due to the high number of metabolites screened for significance.

**Metabolite identification.** Since metabolite identification in mass spectrometry is based on comparison with existing spectral data, it is limited by the lack of chemical standards for a direct identification of all phytochemicals present in the sample, by
insufficient (although constantly expanding) spectral libraries, and by uncertainty related to putative annotations of the metabolites. The uncertainty is further increased when using in silico generated MS/MS reference spectra, which are calculated based on the expected fragmentation mechanisms of known chemical structures in databases, because this often results in too many candidates to correctly determine the annotation of the metabolite [417]. Therefore, the identification step introduces a major limitation and bottleneck in the studies included in this dissertation. The identification step should be carried out with caution, constantly questioning oneself whether the given annotation (e.g. regarding isomeric compounds) is sufficiently supported by the data. In addition, any automated identification processes using any other than in-house databases always need manual curation to confirm the annotations. Any metadata regarding the identifications (exact mass, MS/MS spectra, retention time) should be published along with the level of identification [272] for transparency of the research conducted and to provide additional reference data that can be utilised (and confirmed) by future research. For example, in Study I, luteolin was putatively annotated (level 2) based on a spectral match with the METLIN database. Later, due to expansion of our in-house database, the identity of the metabolite has been confirmed with a reference standard (level 1). Often, any novel biomarkers discovered in nontargeted studies will be first putatively annotated with some uncertainty, and in further studies, a reference standard can (and should be) purchased or synthesised to verify the identification and allow further discussion of the relevance of the biomarker. Many in-house libraries have a relatively comprehensive set of endogenous metabolites as reference standards, but with phytochemicals, it is more challenging to gather a sufficient collection of standards because of their low availability, high number of compounds and the related high expenses to acquiring such a library.

Method standardization and validation. Another issue inherent in nontargeted metabolomics – particularly in LC–MS-based approaches – is the limited validation and nearly complete lack of standardization of the methods. In other words, each platform typically develops their own nontargeted method more or less independently, based on the compound classes or sample matrices most relevant for their research. Thus, some of the data produced by the unique platforms (such as retention times) can only be used as further reference within the same platform, making it more difficult to build public databases that would allow reliable identifications universally. Nevertheless, as a discovery-oriented approach, nontargeted metabolomics does not necessarily need standardization, but knowledge of the limitations (such as analytical coverage) is important to reduce bias when interpreting the results. In the publications included in this doctoral thesis, a wide array of compounds, ranging from extremely hydrophilic (e.g. betaines) to lipophilic (e.g. phosphatidylcholines) and from relatively small to large metabolites (90 to 781 Da) were characterised, indicating that the methodology has a sufficient analytical range at least in terms of lipophilicity and mass range. However, only few lignans were detected in the studies despite of their moderate abundance in whole grains, suggesting that using a single
metabolite extraction method is less suitable for certain compound classes while favouring others, thus leaving at least some potentially important compounds outside of the results. In our recent paper, we proposed a set of standards for the validation of the analytical coverage of nontargeted plant metabolomics [76]. Our aim is to develop and validate a consensus method for the nontargeted analysis of phytochemicals, covering a wide range of compound classes, which could be adopted by new platforms and enable a more efficient data sharing between platforms. However, LC–MS as an analytical technique has limited repeatability compared to e.g. NMR and GC–MS, and therefore the implementation of fully validated and standardised universal methods may be out of reach with the current technology. It would also require a strong consensus within the LC–MS metabolomics community, which is currently lacking.
11 CONCLUSIONS

Whole grains are an abundant source of phytochemicals, which contribute to the health benefits obtained from diets rich in wholemeal cereals and other plant-based foods. The scientific community is only beginning to understand the mechanisms underlying the reduced disease risk and improved cardiovascular and metabolic health associated with whole grain intake. In order to exert their physiological functions \textit{in vivo}, phytochemicals need to be liberated from the cereal bran matrix, metabolised by gut bacteria and/or endogenous mechanisms, absorbed into the circulation – and perhaps most importantly – consumed in the first place within food. The studies included in this dissertation aimed to provide knowledge on how food processing techniques could improve the bioaccessibility and extend the variety of phytochemicals from whole grains and how gut microbial metabolism transforms them prior to absorption.

Among the primary findings in this dissertation was the major impact of enzymatic bioprocessing and sourdough fermentation on the metabolite profile of whole-grain rye, increasing the levels of several potentially beneficial compounds possibly contributing to the health implications of rye bread, which is typically baked using sourdough. The results support the use of traditional sourdough baking, possibly enhanced with enzymatic bioprocessing, in producing functional foods with higher potential for a beneficial impact on health. In addition, we showed that bioprocessing does not greatly reduce the production of microbial metabolites of phenolic acids from rye bran, thus allowing the absorption of both high amounts of phenolic acid precursors from the small intestine and their microbial metabolites from the colon. The results provided suggestions to expand the gut microbial metabolic pathways of phenolic acids.

The major novel finding in this thesis work was the gut microbial origin of several betainized compounds, including 5-aminovaleric acid betaine (5-AVAB) and alanine betaine, many of which were observed earlier in mice consuming a diet enriched with cereal bran. Here, we studied mice receiving bran-enriched diets, compared the betaine levels in the tissues of conventional and germ-free mice, and performed an \textit{in vitro} human colonic fermentation of rye bran to elucidate the role of gut microbiota in betaine metabolism and the influence of bran-enriched diet in the gut microbial composition of mice. The results indicate that microbiota is a main source for the production of betainised compounds, the precursor likely being glycine betaine abundantly available from whole grains. The microbial composition was altered, increasing the abundance of microbial taxa associated with health benefits. Our study introduces a novel group of microbial metabolites with plant origin and biological activity, which are potential mediators of the positive health impact of whole grains. As future directions, the novel group of betainised gut microbial metabolites requires further investigation with targeted and mechanistic studies to determine the metabolic pathways and the specificity of the metabolism among microbiota. The
studies included in this doctoral dissertation provided evidence that bran bioprocessing, while significantly increasing the bioaccessibility of phenolic acids from cereal bran, may in fact reduce the abundance of beneficial gut bacteria. However, the findings were inconclusive for different cereals and require further confirmation. The differential metabolites of sourdough fermentation include several interesting compounds with implications to e.g. modulating the insulin response. The role of these compounds in the rye factor and overall health effects of whole grains merits continued research. LC–MS-based nontargeted metabolomics used in this doctoral dissertation is a powerful method in deciphering the complex metabolic fingerprints of whole grains and other plant-based samples. However, there is room for improvement: currently, typically only 20% of the statistically significant differential metabolites can be identified or putatively annotated because of the low number of reference spectra compared to the vast number of existing plant secondary metabolites. The scientific community should increase resources for analysing more such compounds and providing the results in public databases for reference. In method development, the introduction of standardised nontargeted LC–MS methods validated for an optimal coverage of phytochemicals would promote data sharing, and it could facilitate the metabolite identification for more comprehensive image of the metabolic profile in the studied samples. The molecular and physiological mechanisms of phytochemicals in improving health and preventing disease are still far from being established. The coming decade will likely show if our knowledge on these important molecules increases to the level where their potential can be fully utilized.
REFERENCES

1. Fao, FAOSTAT. 2016, Food and Agriculture Organization of the United Nations.


44. Martinez, I., et al., Gut microbiome composition is linked to whole grain-induced immunological improvements. The ISME journal, 2013. 7(2): p. 269.


144. Verardo, V., et al., *Determination of free flavan-3-ol content in barley (Hordeum vulgare L.) air-classified flours: comparative study of HPLC-DAD/MS and


309. Taylor and G. Francis, Dictionary of Natural Products. 2015.


## APPENDIX 1. SUPPLEMENTARY MATERIAL FOR STUDY IV

Supplementary Table 1. Betainised compounds observed in the current study with their LC–MS identification characteristics and presence in each studied sample type. In case the compound was detected only as trace in a sample, the identification was based on exact m/z and retention time. 1 = identified with a reference standard; 2 = putatively annotated based on publicly available exact m/z and MS/MS spectra.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Observed characteristics in LC–MS (HILIC+)</th>
<th>Identification in samples</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>m/z</td>
<td>∆ppm</td>
</tr>
<tr>
<td>TMAO</td>
<td>76.0758</td>
<td>−5.8</td>
</tr>
<tr>
<td>Glycine betaine</td>
<td>118.0863</td>
<td>−4.2</td>
</tr>
<tr>
<td>Alanine betaine</td>
<td>132.1019</td>
<td>−8.3</td>
</tr>
<tr>
<td>Trigonelline</td>
<td>138.0552</td>
<td>−6.2</td>
</tr>
<tr>
<td>Proline betaine</td>
<td>144.1019</td>
<td>−7.6</td>
</tr>
<tr>
<td>4-aminobutyric acid</td>
<td>146.1173</td>
<td>−9.3</td>
</tr>
<tr>
<td>betaine</td>
<td>158.1177</td>
<td>−6.0</td>
</tr>
<tr>
<td>Hydroxyproline</td>
<td>160.0968</td>
<td>−7.0</td>
</tr>
<tr>
<td>Valine betaine</td>
<td>160.1332</td>
<td>−6.9</td>
</tr>
<tr>
<td>5-AVAB</td>
<td>160.1334</td>
<td>−5.6</td>
</tr>
<tr>
<td>Isoleucine betaine</td>
<td>174.1485</td>
<td>−8.3</td>
</tr>
<tr>
<td>Glutamine betaine</td>
<td>189.1234</td>
<td>−5.6</td>
</tr>
<tr>
<td>Trimethyllysine</td>
<td>189.1600</td>
<td>−4.5</td>
</tr>
<tr>
<td>Glutamic acid</td>
<td>190.1073</td>
<td>−6.2</td>
</tr>
<tr>
<td>Histidine betaine</td>
<td>198.1237</td>
<td>−5.6</td>
</tr>
</tbody>
</table>
Phenylalanine 208.1332  −5.3  1.29  20 V: 107.048 (100), 131.048 (80), 103.054 (62), 58.066 (50), 208.133 (46; [M + H]+), 79.053 (48)
betaine

Tryptophan 247.1467  6.1  1.32  20 V: 146.061 (100), 188.072 (54), 60.082 (48)
betaine

Supplementary Figure 1. Principal component analysis (PCA) and redundancy analysis (RDA) of the microbial composition (relative abundance data) from the caecal contents of the C57BL/6J mice, with the control diet groups (left) and treatment groups (right) treated separately. The RDA illustrations display 10 best-fitting microbial genera.
Supplementary Figure 2. Box plots of selected bacterial genera in the caecal contents of the C57BL/6J mice with significantly different relative abundances in Kruskal–Wallis one-way ANOVA between all diet groups. Outliers have been marked with a circle and extreme outliers with an asterisk.
Supplementary Figure 3. The average abundance (as signal counts) of all the identified betainized compounds in the tissue samples of GF and MPF mice. The error bars signify an error of 1 SD. Asterisks based on Mann–Whitney U test between the groups: * p < 0.05; ** p < 0.01; *** p < 0.001.
Supplementary Figure 4. The levels (as signal counts) of betainized compounds detected with UHPLC–qTOF-MS in 10 different wheat and rye samples, including processed (white or sifted, containing only the endosperm) and whole-grain flour (containing all the edible parts of the grains in their original proportions), processed and whole-grain breads, and sourdough fermented breads tailor-made at VTT Technical Research Centre of Finland. The error bars signify an error of 1 SD.

Supplementary Figure 5. The relative abundance of bacterial phyla in the caecal contents of the studied C57BL/6J mice, divided into control diet groups and those fed with the bran-enriched diets. The samples are arranged in ascending order of the most abundant phylum, Firmicutes, which ranges from 52% to 93% of the microbial population.
Whole grains are an abundant source of phytochemicals, which contribute to the health benefits of plant-based foods. This study showed that enzymatic bioprocessing, sourdough fermentation, and colonic microbial metabolism have a major impact on the abundance of whole grain phytochemicals, affecting their bioavailability and even producing novel metabolites, including betainized compounds. The study further supports the applicability of metabolomics in food science and nutrition.