LC-MS based Metabolomics: Biomarker Discover and Bioinformatics Tools

Tu Hu[1]

Sep, 2019

[1]Supervised by Lars Ove Dragsted and Gözde Gürdeniz

Contents

Chapter 1

Introduction

1.1 Biomarkers of Food Intake (BFIs)

Definition of Biomarker

Biomarker Classification

Biomarker Systematic review

Credibility of BFIs: Putative biomarker, Candidate biomarker, partially validated biomarker and fully-validated biomarker

Biomarker Validation

1.2 Discovery of BFIs

Chapter 2

A Mini-review on Biomarkers of Whole Grain Wheat and Barley Intake

2.1 Abstract

Due to lacking objective food exposure measurement, epidemiologic studies showed mixed results on whole grains' health beneficial effects. Meanwhile, increasing evidence showed each whole grain type could benefit health differently. In order to accurately quantify whole grain exposure, there is a demand to discover intake biomarkers for each whole grain type. This mini review referred the systematic literature review guidelines and examined biomarkers for whole grain barley and whole grain wheat intake. For whole grain barley, none intake biomarker has been reported. For whole grain wheat, combining total alkylresorcinol and homologous ratio of C17:0/C21:0 was proposed as biomarkers to measure whole grain wheat intake. Other promising intake biomarkers were summarized as well.

2.2 Introduction

Whole grains (WGs) contain a lot of non-nutrient compounds in the bran. These non-nutrient compounds might benefit health.

Several dietary guidelines suggested WG intake[1]. A recent meta-analysis confirmed high intake of WGs is associated with reduced risk of cardiovascular disease, cancer, and all cause and cause specific mortality[2]. Recommendations for WG intake have often been unclear or inconsistent with regard to the amount and types of whole grain foods that should be consumed to reduce chronic disease and risk of mortality.

Increasing evidence showed different WG types (such as wheat, rye, oat,

barley etc.) could benefit health differently. However, classical self-reported measurement tools (e.g. food diaries and food frequency questionnaires) used in observational studies could cause biases and confoundings to distinguish each cereal type.

Therefore, discovering BFIs of each whole grain type could potentially provide a tool to accurately quantify their exposures. WGs' health beneficial effects could be further elucidated.

This mini-review aimed at systematically examining available literatures to obtain information of potential biomarkers for WG barley and wheat intake. This will prioritize further identification and validation of the thesis work.

2.3 Materials and Methods

This review referred the systematic BFIRev methodology[3]. The flowchart was included in Appendix (Fig-2.2 and Fig-2.3)

The objective of this literature review was to identify and evaluate reported potential biomarkers for dietary assessment for whole grain wheat and whole grain barley.

Keywords as suggested in the guidelines[3] were used to search in 3 database (PubMed, Web of Science, Scopus). Keywords used for searing BFI barley in human: (barley) AND (biomarker* OR marker* OR metabolite* OR biokinetics OR biotransformation OR pharmacokinetics) AND (intake OR meal OR diet OR ingestion OR consumption OR eating OR food) AND (human* OR men OR women OR patient* OR volunteer* OR participant*) AND (trail* or experiment OR study) AND (urine OR plasma OR blood OR serum OR excretion OR hair OR toenail OR faeces OR faecal water). The first element was changed to wheat for wheat intake biomarker searching.

Due to limited amount of searching results, barley searching scope was expanded to animal studies. Therefore, the keyword (animal* OR goat OR sheep OR cow OR mice OR mouse* OR animal model* OR dog*) was used to replace the previous 'human*' entry. In addition, 'feed' was added to 'food' entry.

Other database including HMDB[4], FoodDB[5], PhenolExplorer[6], Dictionary of Food Compound[7] were also used to explore compounds present exclusively in WG barley and wheat.

In order to verify the uniqueness of compound, the same keywords combinations were used but with compound name instead of 'wheat' and 'barley'.

2.4 Results

2.4.1 WG barley

The literature search got 129 records after removing duplicates from merged 3 database results. Within them, none of the studies directly investigated WG barley intake biomarkers. This could be due to limited barley consumption

in population. Although barley is the 4th most produced cereal grains worldwidely, they are mostly used for brewing or feed. Approximately only 4% is directly consumed[8].

When the scope expanded to animal studies, the search results still did not show any direct research about BFIs. Most of animal studies were interested in how barley feed can benefit the growth of animals or quality improvement of animal-source products[9, 10].

A 2-month intervention study[11] incorporated 75% refined drum wheat and 25% WG barley. The fecal samples showed significant change in microbiota and metabolome after intervention[11]. However, no specific metabolite can indicate WG barley intake.

ARs and their metabolites may not indicate WG barley intake. Several observation studies[12, 13] investigated correlation between ARs metabolites and whole grain intake. Although these studies tried to cover more whole grain species, for example, one study[13] listed 7 types of commonly consumed WGs in American populations in the Food Frequency Questionnaire (FFQ)[1], barley was not solely listed. Therefore, although ARs and their metabolites got good correlation with these 'Whole-grain intake'. Readers should be cautious to apply these markers to WG barley intake. In addition, ARs concentration in cereal barley is much lower compared with WG wheat and rye, with similar concentration with refined wheat and rye flours (Table-2.1).

Cereal	Conc. range in cereal	Conc. average or range in WG flour	Conc. average in refined flour	Main homologues	C17:C21 homologues ratio
Rye	360-3200	972	90	C17, C19, C21	0.8-0.9
Wheat	761-8390	490-710	36	C19, C21	0.07-0.1
Barley	55.8-98.2	NA	NA	C19, C21, C25	NA

Table 2.1: Prensence of ARs in Cereal Grains, adapted from [14–16](unit: µg/g dm), conc. varies due to different species and milling methods.

Most search results focused on barley's *effect biomarkers* as defined by Dragsted[17] and Gao[18], such as bowel health indicators[19], postprandial glucose and insulin response[20], lipid profiles and cardiovascular diseases (CVD) markers[21], etc. However, in these intervention studies, compliance monitoring lacked objective markers.

2.4.2 WG wheat

Overview

The literature search got 312 results after removing duplicates from merged results. Some articles were found from the references of searched results. One intervention study and two observation studies were included in Table-**2.3**.

[1]Dark breads, High-fiber or bran cereals, Cooked cereals and grits, Regular granola, Granola bars and cereal bars, Plain popcorn (no butter) or low-fat microwave popcorn, Buttered or gular microwave popcorn

Surprisingly very few studies investigated WG wheat intake biomarkers although WGs seem to be the hotspot area in food and nutrition research.

Alkylresorcinols (ARs) and Homologous Ratio C17:0/C21:0

Total Alkylresorcinols (ARs) can indicate WG wheat and rye intake, meanwhile AR homologous ratio C17:0/C21:0 can indicate relative compositions of WG wheat and rye in the diet.

Within commonly consumed plant-based food, alkylresorcinols present high concentration exclusively in bran part of wheat and rye. AR homologous ratio C17:0/C21:0 was first reported by cereal scientists in 2004 to distinguish WG rye and wheat cereal [22]. In grains, rye has homologous C17:0/C21:0 ratio close to 1.0, while wheat around 0.1, durum wheat around 0.01.

Food items	No. subjects	Study design	Sample type	Analytical method	Candidate biomarker(s)	Identifier	Reference
WGs[2]	2845	Observation (11 countries)	fasting and non-fasting plasma	GC-MS	AR C17:0/C21:0	HMDB0038530 HMDB0031035	[23]
WG wheat WG rye	73	Observation (after 0.1-3 years)	plasma	GC-MS	Total ARs (C17:0,C19:0, C21:0,C23:0,C25:0) Ratio of AR (C17:0/C21:0)	HMDB0038530 HMDB0030956 HMDB0031035 HMDB0038524 HMDB0038485	[24]
WG wheat WG rye	39	Intervention (cross-over)	plasma	GC-MS	Ratio of AR C17:0/C21:0	HMDB0038530 HMDB0031035	[25]
WG wheat WG rye	15	Intervention (cross-over)	plasma serum enterolactone	GC-MS	Ratio of AR C17:0/C21:0	HMDB0038530 HMDB0031035	[26]

Table 2.2: Biomarkers of Wheat Intake Reported in Intervention study

Further this marker was proposed by nutritionists to distinguish WG rye and wheat intake. In 2005, Linko[25] first investigated this biomarker in human plasma to measure food exposure. The intervention study showed the potential of this marker AR C17:0/C21:0 to distinguish WG wheat and rye in diet in healthy postmenopausal women. For rye-dominated diet, the ratio was 0.84 and for WG wheat-dominated diet, the ratio was around 0.53. Further in 2007, Linko-Parvinen[3] validated this marker in healthy adults by an intervention study [26]. In plasma, the value was 0.1 after WG wheat intake, 0.6 after WG rye intake. In erythrocytes, the value was 0.06, 0.33 respectively after WG wheat and rye intake. This study also implied ARs could be transported in human plasma lipoproteins.

However, the AR homologues ratio C17:0/C21:0 was unable to differentiate WG diet and refined cereal diet as reported by Landsberg[27]. But WG diet and refined diet can be distinguished by total ARs concentration in plasma.

EPIC[4] cohort study [23]further validated this marker in 2014. This observation study investigated plasma ARs and the C17:0/C21:0 ratio of subjects from 10 European countries. The result showed that Greek, Italian, Dutch and UK

[3]Probably, she married and changed last name. (This footnote will be deleted in final thesis)

[4]European Prospective Investigation into Cancer and Nutrition

participants of whom the diet was dominated by wheat, had low C17:0/C21:0 ratio in plasma. Whereas Danish, German and Swedish subjects had high C17:0/C21:0 ratio. French and Norwegian subjects had intermediate ratio. This marker showed reverse correlation with WG wheat consumption in the population.

Applications in Type II Diabetes Research

Combining plasma total ARs concentration and the ratio (C17:0/C21:0) can subjectively estimate WG wheat and rye intake and approximate their relative compositions in the diet. This will provide information of specific cereal species intake amount and facilitate their disease preventative effects study.

For example, an observation study showed, in Chinese population AR metabolite DHPPA was correlated with lower odds of type II diabetes and impaired glucose regulation[28]. However, DHPPA as AR metabolite can come from either whole grain wheat or rye, therefore not specific to WG wheat.

Other two researches can better show that, actually WG rye could be more favorable for type II diabetes. An observational study showed in a population with metabolic syndrome, plasma AR C17:0/C21:0 was associated with increased insulin sensitivity[29]. Further, it was observed that in healthy Scandinavian populations, plasma total ARs concentration was not correlated with type II diabetes risk. However higher C17:0/C21:0 ratio (implicating more rye intake) was associated with increased insulin sensitivity[30]. These results imply that, a whole grain diet dominated by rye could be favourable for type II diabetes prevention.

In EPIC cohort study[23], there is an interesting phenomena. Rye has higher constitutions in Danish populations's WG source (70%) than Swedish (55%) on average. Regarding C17:0/C21:0 ratio, Danish participants in EPIC cohort showed lower value (0.37) than Swedish participants (0.43). However, Swedish participants were healthy adults, while participants from Denmark in EPIC study were obese or over-weights subjects. Those participants may have different dietary habits and consume less rye than average Danes. This may imply rye could also be favorable in weight control.

These studies indicate that, application of biomarkers for sub-type cereal intake could provide more information to study each cereal subtype's health beneficial effects.

Other potential markers

Searching results also showed some *Food compound intake biomarkers (FCIBs)* research as defined by Gao[18] such as phenolic compounds[31], benzoxazinoids (BXOs)[32, 33], phytoestrogen[34], phytosterol and lignan[35] and *effect markers* Dragsted[17] microbioal derivitives[33]. These compounds also present in other food, not specific for WG wheat. These results were summarized in Appendix.

In this article[35], author performed an intervention study using refined wheat bread as control group, authors proposed a panel of metabolites con-

sisting 7 AR metabolites, 5 BXO metabolites and 5 phenolic acid derivatives can be used to objectively monitor WG wheat intake. The concentration of these phytochemicals varied in different cereal grains. Therefore, a combination of their metabolites could potentially indicate intake of different cereals. However, this conclusion needs to be further identified in other intervention and observation studies.

2.5 Conclusions

WG barley attracted a lot of interest due to its health beneficial effects for preventing chronic disease. However, because barley is seldom consumed directly, currently there's no biomarkers reported for WG barley intake both in human and animal studies. A lot of sparse information was reported from cereal and food chemistry could give hints for identification and validations of WG barley's intake biomarkers.

Total ARs and their metabolites were reported to potentially indicate WG wheat and rye intake. The homologues ratio of ARs C17:0/C21:0 was proposed to distinguish which whole grain type dominates in the diet.

2.6 Discussions

2.6.1 Potential BFIs for WG barley

Search results in food chemistry, cereal science and plant science database showed some compounds exclusively present in barley. These could give hints for further identification. The results were summarized in Table-**2.4** in **Appendix**.

2.6.2 Exclusion Criteria of WG Wheat intake biomarker

In order to identify specific marker to distinguish WG wheat intake from refined wheat or other WG types, following exclusion criteria was used:

1. The terms "WGs or dietary fibers or cereal fibers or bread" were used, however no details about how much WG wheat was included.
2. In application studies[5], intake amount was only measured by biomarkers but without any other validation (such as FFQ, FR etc)
3. Intervention was WG diet or meal. Markers were not specific to wheat.

Within searched results, most intervention studies used WG diet containing several types of cereals as a comparison with refined diet. The reason could be, most people consume more than one cereal type. Therefore, research interest were also in the combination of several WG cereal types.

[5]"application study" refers, applying BFIs to estimate correlation between WG wheat intake and disease risk and frequencies

Another excluded example is ambiguity of cereal types[33], subjects were classified as consumers of "none-bread", "white bread" or "whole grain bread". However, no further details about cereal types of the WG bread.

In observational studies, estimations of food exposure are majorly self-report based. In these self-based surveys, participants had difficulty recalling and distinguishing the different cereal species. Results were normally not detailed to each WG type. Therefore, it is difficult to assign the biomarker to specific WG. Food frequency questionnaire causes high deviations distingushing each sub-type cereal by its nature.

Those non-specific markers were listed in appendix.

2.6.3 AR metabolites for WG wheat intake and Geographical Applicabilities

Biomarkers for WG wheat intake could have their geographical applicabilities. Here, "geography" refers to an area having similarly available food resources.

In the case of WG wheat intake, some biomarkers were reported to be capable of distinguishing whole grain wheat intake in US population and proposed as putative intake biomarkers, including DHBA, DHPPA, DHBA glycine, 3,5-DHPPTA[36], ARs (C19:0, C21:0, C23:0)[37]. However, according to current criterial of BFIRev[3] and validation[38], these biomarkers were not classified as "candidate biomarker" because of their non-specificities:

- DHBA, DHPPA, DHBA glycine, 3,5-DHPPTA: They are AR metabolites (Fig-2.1) and detected in urine. When ARs are metabolized, information of their homologue ratio can not be traced. Therefore, it's impossible to distinguish their origins are WG wheat or WG rye.
- ARs (C19:0, C21:0, C23:0): WG wheat mainly consists these three homologues. They were reported well correlated with WG wheat intake in an intervention study[37]. However, WG rye also consists these metabolites.

However, to our best knowledges, rye is rarely consumed by US population. Therefore, in such a "geography" where WG wheat is the only AR source. These markers have their geographical applicabilities although are not be world-widely applicable.

This also raised another question about biomarker validation. Regarding confidence level of biomarkers, when a "putative" biomarker can be upgraded to "candidate" biomarker, it needs to be confirmed by more trails and preferably by another design or in a different population. Therefore, this may not fit for these "geographically-applicable" biomarkers.

2.6.4 Analytical platform of ARs

Currently, in vast majority studies, ARs were discovered and quantified in GC-MS. Whether phase II metabolites may interfere quantification results is unknown. This could be investigated more to improve this markers' analytical stabilities.

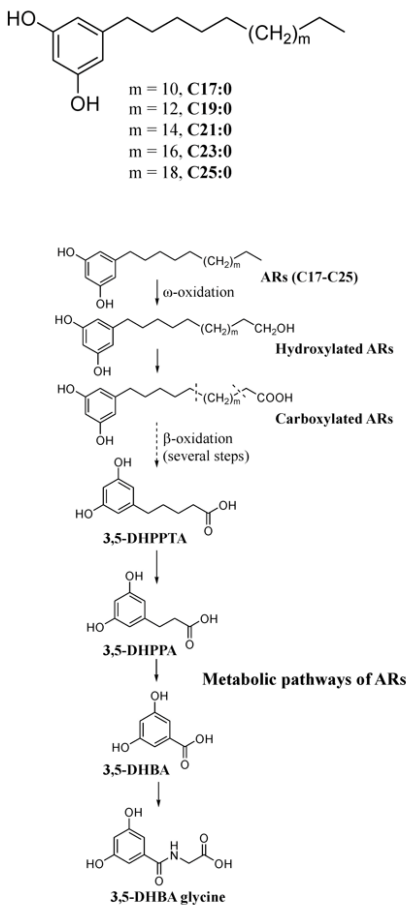

Figure 2.1: Structure of ARs and suggested metabolic pathway, adapted from [39]

Food items	No. subjects	Study design	Sample type	Analytical method	Candidate biomarker(s)	Identifier	Reference
WG wheat bread Refined wheat bread	12	Intervention	24-h urine	HPLC-CE[6]	DHBA DHPPA DHBA glycine 3,5-DHPPTA	HMDB0013677 HMDB0125533 InChi: QVGDKHUNWDVPOR-UHFFFAOYSA-N InChi: QHXNJIMVPAFCPR-UHFFFAOYSA-N	[36]
WG wheat (3 or 6 servings)	19	Intervention (crossover 1 week)	fasting plasma	GC-MS	ARs (C19:0,C21:0,C23:0)	HMDB0030956 HMDB0031035 HMDB0038524	[37]

Table 2.3: "Geographically applicable (specific)" Biomarkers of Whole Grain Wheat Intake

2.6.5 Necessities and Challenges of Discovering Biomarkers for WG Wheat Intake

In order to clarify health beneficial effect of each cereal subtype, it is important to accurately quantify exposure amount of each sub-type cereal. BFIs showed their strengths in studying WGs. Discovering intake biomarker for each subtype cereal grain becomes necessary. Currently, most studies showed interest in WG effect biomarkers.

One of the challenges in BFIs discovery of WG is chemical compositions of most of WGs were not systematically studied. Food compound, natural product, phytochemical database also need to be expanded.

2.7 Appendix

No	Candidate biomarker	Formula	Chemical group	Presence in Food	Reference
1	Hordenine	$C_{10}H_{15}NO$	alkaloid	germinating barley, beer and other plants	[40]
4	Hordatine A	$C_{28}H_{38}N_8O_5$	alkaloid	only reported in barley	FoodDB (002330)
4	Hordatine B	$C_{29}H_{40}N_8O_5$	alkaloid	only reported in barley	FoodDB (002328)
2	Distichonic acid A	$C_{10}H_{18}N_2O_8$	gamma amino acids and derivatives	only reported in barley	FoodDB (18164)
3	Distichonic acid B	$C_{10}H_{18}N_2O_8$	gamma amino acids and derivatives	only reported in barley	FoodDB (018165)
5	14,16-Nona cosanedione	$C_{29}H_{56}O_2$	ketone	only reported in barley	FoodDB (013891)
6	N-Norgramine	$C_{10}H_{12}N_2$	indole	only reported in barley	FoodDB (017815)

Table 2.4: Potential Biomarkers for WG Barley Intake

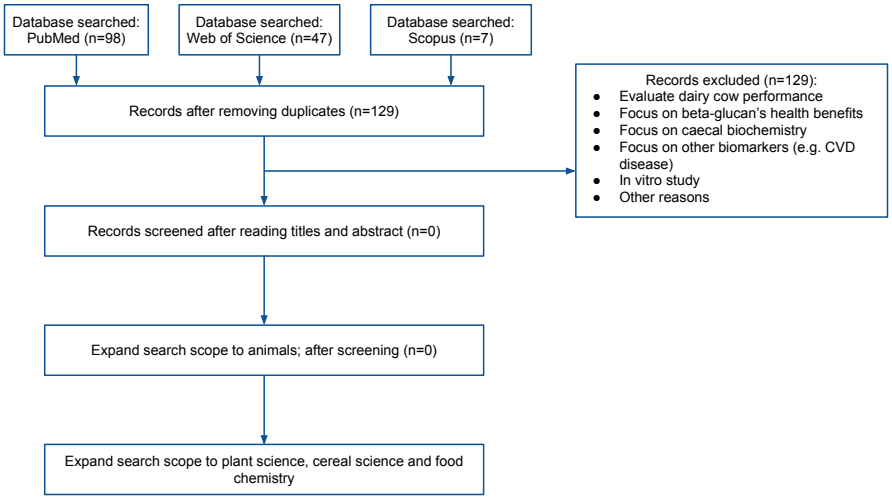

Figure 2.2: Flow Chart of Literature Searching and Screening for Articles of WG Barley Intake Biomarkers

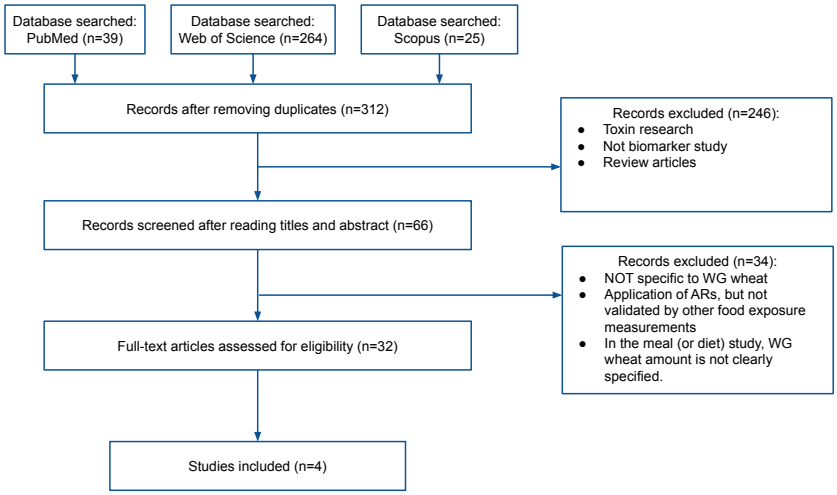

Figure 2.3: Flow Chart of Literature Searching and Screening for Articles of WG Wheat Intake Biomarkers

Dietary factor	No. subjects	Study design	Sample type	Analytical method	Candidate biomarker(s)	Reference
Wheat bran, Wheat aleurone	14+13	randomized, cross-over, intervention	plasma	LC-MS/MS (Microbiology assay for folate)	betaine choline folate dimethylglycine (DMG)	[41]
None-bread, White bread, WG bread	155	observation[7]	urine	HPLC-qTOF-MS	Benzoxazinoid-related metabolites (HHPAA, HBOA glycoside) ARs-related metabolites (DHPPA glucuronide, DHPPTA sulphate microbial-derived metabolites	[33]

Table 2.5: Reported markers distinguishing WG wheat intake, but NOT specific

Chapter 3

Discovering Barley Intake Biomarkers in Urine by UPLC-MS Based Untargeted Metabolomics

Chapter 4

Barley Intake Biomarker: Compound Identification and Structure Elucidation

4.1 Abstract

4.2 Introduction

4.2.1 Phytosterol: Structure, Post-harvest Degradation and Fragmentation Behaviour

Phytosterols ubiquitously occur in plant-based food[42]. They were claimed to have health beneficial effects, such as lowering cholesterol. They occur in food as free sterols (FS), steryl esters (SE), and glycosylated conjugates comprised of steryl glucosides (SG) and acylated steryl glucosides (ASG).

Specific sterol profiles characteristic to certain plant families have been identified showing that a broad range of minor sterols occurs as free sterols or glycosylated conjugates

This ion could be stanol (a sub-type of phytosterol with a saturated B-ring) derivative inferred from its C-ring fragmentation behaviour:

- higher intensities of m/z 149 than both 147 and 145
- higher intensities of m/z 161 than both 159 and 163

4.2.2 β-glucuronidase for metabolites analysis

Glucuronidation in METABOLISM

Glucuronidation, conjugation with glucuronic acid, by the human UDP-glucuronosyltransferase (UGT) family of enzymes plays an important role in the metabolic fate of many

drugs and other xenobiotics. This biosynthetic reaction also has a role in the conjugation and excretion of endogenous substrates, such as steroids, bilirubin, and bile acids.9 UGT activity results in the conjugation of glucuronic acid to substrates containing sulfhydryl, hydroxyl, aromatic amino, or carboxylic acid moieties. The glucuronides formed are more polar (water soluble) than the parent organic substrate and are generally excreted through the kidney.

glucuronide group hurders the structure elucidation of metabolites. therefore, in structure elucadation stage, in order to confirm the structre. it's recommended to hydrolyze the glucuroide group.

Although there are a lot of studies trying to elucidate sturucture of phase II metabolites (i.e. glucuronate), the exact conjugation site can only be determined by additional NMR analysis.

4.3 Materials and methods

4.3.1 Chemicals

Sitostanol standard (CAS Number 83-45-4, Avanti Polar Lipids Inc., USA) was transported and stored in -20 °C. Ethanol

Beta-glucuronidase (CAS Number 9001-45-0, E.C. number 3.2.1.31, Sigma-Aldrich, from E. Coil, optimal pH 6-7)

4.3.2 Apparatus

UPLC-MS system (column C18, QTOF (VION, Waters, Milford, USA), Water Bath

4.3.3 UPLC-MS/MS analysis of Sitostanol

4.3.4 Deconjugation experiments

1. Prepare phosphate buffer[1] (0.1 M, pH=6.8, 50 mL):
   - Prepare 40 mL of distilled water in a volumetric bottle.
   - Add 0.656 g of monosodium phosphate to the solution.
   - Add 0.352 g of disodium phosphate to the solution.
   - Adjust pH using HCl or NaOH.
   - Add distilled water until volume is 0.05 L.
2. Prepare enzyme solution (4 mg/mL): dissolve 0.0060 g β-Glucuronidase in 1.5 mL phosphate buffer. Store in -20°C freezer[2]

[1] Calculated by AAT Bioquest

[2] Sigma Product Information: A solution in 75 mM phosphate buffer, pH 6.8, >5 mg/ml may be stored at -20 °C for up to 2 months with little or no loss of activity.

3. Prepare urine samples: thaw samples in the fridge and centrifuge (3000 rpm, 2 min). Prepare 2 Eppendor tubes, one labeled as 'blank', one as 'treatment'. Transfer 100 µl to each Eppendorf tube.

4. Enzymatic hydrolysis reaction: add 50 µL phosphate solution to 'blank', 50 µL enzyme solution to eppendorf tube, incubate in 37 °C for 1.5 h.

5. Denature enzymes to terminate the reaction

    - Add 50 µL MeOH to the solution and vortex mix for 1 min
    - Centrifuge at 3000 rmp for 3 min
    - Add solvent A 300 µL
    - Transfer supernatant to vials fur further LC-MS/MS analysis.

4.4 Results

4.4.1 Retention Time (RT) and m/z in Different Matrix

Matrix	RT	m/z (ESI+)	Annotation
Whole Grain Barley	6.88	291.2683	Unknown
Urine	6.71	291.2683	Unknown
Standard	8.60	399.3989	[Sitostanol-H2O+H]-

4.4.2 Enzymatic Deglucuronidation

Both ions were hydrolyzed by enzymes. In control group, ions have much higher intensities than treatment group. However, their deglucuronited ions were not all detected. ion 341.2675 was detected in RT 0.88 indicating it is a highly polar compound.

4.4.3

Chapter 5

Discovering Novel Intake Biomarkers of Whole Grain Wheat Intake by LC-MS Based Untargeted Metabolomics

5.1 Introduction

5.2 Materials and Methods

5.2.1 In vitro glucurnadition mimic experiments

5.2.2 Glucurnadase experiments

5.2.3 logP value prediction

5.2.4 MS/MS

5.3 Results

5.3.1 Summary of Identification

N metabolites were identified. Within them, Ni as level I, Nii as level II.

5.3.2 Predicted logP value of AR metabolites

The general pattern for metabolites are eluting late when chain length increases.

No.	m/z	RT (Quad)	RT (Bi)	MS/MS	Annotation	Suggested Compound
	263.016	5.26	3.31			
	329.0582	2.54	0.97		$[M-H]^-$	3,5-DHBA glucuronate
	329.0582	3.51	1.12		$[M-H]^-$	3,5-DHBA glucuronate
	253.9750	3.58	1.37		$[M-H]^-$	3,5-DHBA sulfate
	210.0411		1.48		$[M-H]^-$	3,5-DHBA glycine
	153.0187	1.18	1.88	109.029	$[M-H]^-$	3,5-DHBA
	357.0898	4.24	1.91		$[M-H]^-$	3,5-DHPPA glucuronate
	715.1716	4.24	1.91		$[2M-H]^-$	dimer of 3,5-DHPPA glucuronate
	261.0077	4.32	2.10		$[2M-H]^-$	3,5-DHPPA sulfate

	ChemDraw			ClogP		
	M	$[M+GluA+H]^+$	$[M+SO3+H]^+$	M	$[M+GluA+H]^+$	$[M+SO3+H]^+$
3,5-DHPPTA	2.01	0.75	0.55	1.48	-0.44	0.10
3,5-DHPPA	1.17	-0.09	-0.36	0.57	-1.34	-0.81
3,5-DHBA	0.81	-0.45	-1.00	0.99	-1.07	-0.54
3,5-DHBA glycine	-0.34	-1.60	-2.34	-0.24	-2.29	-1.75

5.3.3 *in-vitro* glucuronadation of 3,5-DHBA

3,5-DHBA has 3 active sites to be glucuronated. The products could be mono-, di- or tri- glucuronate (Figure-5.1).

RT	m/z	Annotation
2.78	329.05	$[3,5\text{-DHBA} + \text{GluA} - \text{H}]^-$ (glucuronated at carboxylic acid hydroxyl group)
3.97	329.05	$[3,5\text{-DHBA} + \text{GluA} - \text{H}]^-$ (glucuronated at benzene hydroxyl group)
3.6	252.04	$[3,5\text{-DHBA} + 2\text{GluA} - 2\text{H}]^{2-}$
7.34	339.55	$[3,5\text{-DHBA} + 3\text{GluA} - 2\text{H}]^{2-}$

3,5-DHBA monoglucuronate

3,5-DHBA mono-glucuronate exists 2 theoretical possibilities. Because position 3 and 5 are equal. The difference is phenol group or hydroxyl group sitting on carboxyl group be substituted.

Fragment 109 and 137 are specific for glucuronation reaction happening in carboxylic acid or beneze ring. Based on their predicted ClogP value and fragment 109 and 137 ratio. Finally, we confirmed RT 2.78 should be 3-glucuronate-5-hydroxyl-benzoic acid, RT 3.97 should be A. something in between, we have no idea.

3,5-DHBA diglucuronate and triglucuronate

3,5-DHBA di- and tri- glucuronates were also detected. 3,5-DHBA di-glucuronate has 2 possible isomers. But they were not confirmed here.

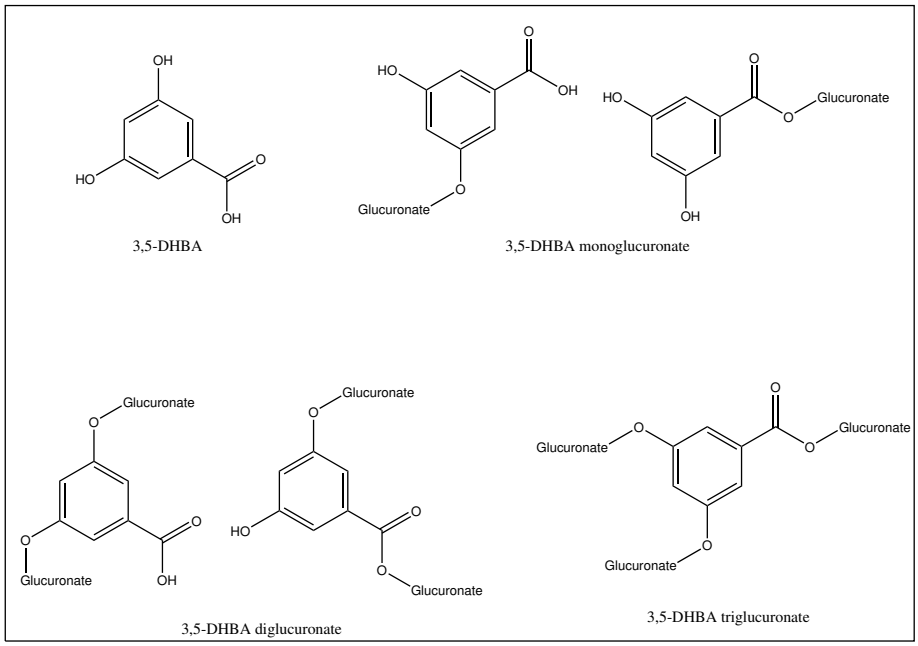

Figure 5.1: 3,5-DHBA

RT	Intensity of 109	Intensity of 137	Ratio of 109:137
2.78	195	1.58e3	1.23e-1
3.08	1.48e3	27.8	52.23
3.97	2.79e4	34.7	804

Table 5.1: 3,5-DHBA monoglucuronate

3,5-DHBA tri-glucuronate does not have isomers. It is detected as [M+3 GluA-2H]2-.

5.3.4 3,5-DHBA sulfate

5.3.5 Identifications of alkylresorcinol (AR) metabolites

5.4 Conclusion

5.5 Discussion

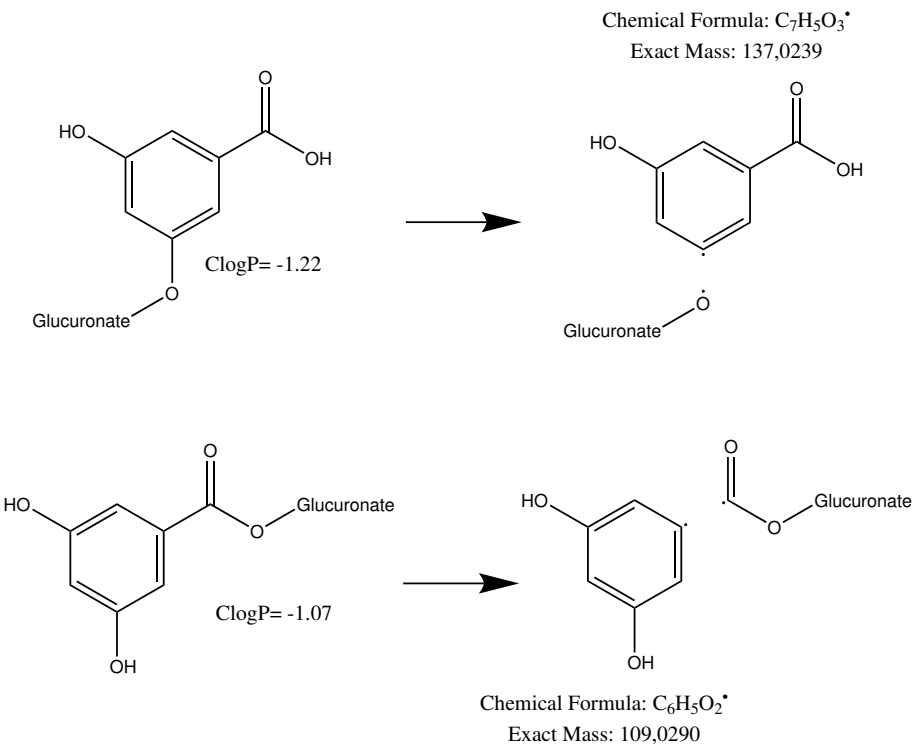

Figure 5.2: Fragmentation of 3,5-DHBA monoglucuronate

Chapter 6

Identification Strategies for BFIs Discovery

In this master thesis work, most of my time was spent on identifying metabolites from biospecies such as urine and plasma. This is same in all biomarker discovery studies. however, this step is very time-consuming. I summaried the identification strategies i used. This might be inspiring for other researchers.

In metabolimics pipeline, identification comes after statistics. Selected features (with retention time and mass-to-charge ratio) should be identified.

step one: database search Normally, 1st step starts from searching database. the database could be a public one or in-house one. Normally in-house database contains RT, fragments, adducts, etc. This might easily match get a level-one identification. however, searching in a public database normally can get extensive hints, e.g. C6H10O might have a lot of hints.

step two: use reference compound 1st, reference compound is available. then purchase them, e.g. in my case, sitostanol did not fly in our Quad method. then, I checked another method, Linda's method to make it fly.

In another case, interesting compound could be an metabolite of the reference compound. Then, we need to some in-vitro metabolism experiments, e.g. oxidation, glucuronadation and sulfazation. If they match

second method is, use glucuronadase or sulfatease to treat the sample and analyze them again, to check whether RT, m/z and MSMS spectra match.

Another challenge is reference compound is not always available. For example in Muyao's PhD thesis, she synthesized her own compound and made a level-1 identification for spinach intake.

6.1 Discussion

6.1.1 Searching in a cohort study

not all metabolites are identifiable. in order to prioritize the identification work, there's one step suggested before identification. That is, check whether this metabolite can also distinguish this intake in another large cohort study. Normally this is the later step of BFIs discovery study, i.e. validation of metabolites. However, these cohort studies normally do not share their data. Also because different analytical platform, data format was used making it difficult to compare.

6.1.2 Challenges and Limitations

Not enough. normally they do not have bio-informatics ideas.

Chapter 7

Discussions

I do not know which markers are better. If another dataset can give me. I think it's gonna be better.