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 by Metabolomics

Currently, there is no systematic review/pipeline defining the BFIs discovery.

why metabolomics is the powerful way to discover novel BFIs? theoritically, food intake will induce metabolome change. However, human physiology, as a s

Chapter 2

Materials and Methods

2.1 Digital Literature Search and Systematic Review

2.2 Multi-variable Data Analysis

2.3 LC-MS Based Metabolome Profiling Systems

In this thesis, two LC-MS analytical systems were used.

2.3.1 MSMS

2.4 title

Chapter 3

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

3.1 Abstract

Due to lacking objective food exposure measurement tools, 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.

3.2 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.

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.

3.3 Materials and Methods

This review referred the BFIRev methodology[3]. The flowchart was included in Appendix (Fig-3.2 and Fig-3.3)

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'.

3.4 Results

3.4.1 WG barley

The literature search retrieved 129 records after removing duplicates. None biomarker of barley intake has been reported from either human or animal studies.

The term "biomarker" mentioned in retrieved results mostly referred to barley intake's *effect biomarkers* as defined by Dragsted[8] and Gao[9], such as bowel health indicators[10], postprandial glucose and insulin response[11], lipid profiles and cardiovascular diseases (CVD) markers[12], etc. For animal studies, "biomarker" mostly referred to the growth of animals or quality improvement of animal-source products[13, 14], which could also be regarded as *effect biomarkers* of animals. However, in these intervention studies, compliance monitoring lacked objective markers.

3.4.2 WG wheat

Overview

The literature search retrieved 312 results after removing duplicates. Some articles were found from the references of searched results. The final result(Table-**3.2**) included one intervention study and two observation studies. Surprisingly to us, very few studies investigated WG wheat intake biomarkers although WGs seem to be the hotspot in food and nutrition research. The reason was described in discussion.

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

Combining total Alkylresorcinols (ARs) and AR homologous ratio C17:0/C21:0 can potentially be used as a biomarker for WG wheat intake.

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 grains [15]. 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[1]	2845	Observation (11 countries)	fasting and non-fasting plasma	GC-MS	AR C17:0/C21:0	HMDB0038530 HMDB0031035	[16]
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	[17]
WG wheat WG rye	39	Intervention (cross-over)	plasma	GC-MS	Ratio of AR C17:0/C21:0	HMDB0038530 HMDB0031035	[18]
WG wheat WG rye	15	Intervention (cross-over)	plasma serum enterolactone	GC-MS	Ratio of AR C17:0/C21:0	HMDB0038530 HMDB0031035	[19]

Table 3.1: 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[18] 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[2] validated this marker in healthy adults by an intervention study [19]. 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[20]. But WG diet and refined diet can be distinguished by total ARs concentration in plasma.

EPIC[3] cohort study [16]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 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[21]. 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, WG rye could be more favourable for type II diabetes prevention. An observational study showed in a population with metabolic syndrome, plasma AR C17:0/C21:0 was associated with increased insulin sensitivity[22]. 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[23]. These results imply that, a whole grain diet dominated by rye could be favourable for type II diabetes prevention.

In EPIC cohort study[16], 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

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

[3]European Prospective Investigation into Cancer and Nutrition

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[9] such as phenolic compounds[24], benzoxazinoids (BXOs)[25, 26], phytoestrogen[27], phytosterol and lignan[28] and *effect markers* Dragsted[8] microbioal derivitives[26]. These compounds also present in other food, not specific for WG wheat. These results were summarized in Appendix.

In this article[28], author performed an intervention study using refined wheat bread as control group, authors proposed a panel of metabolites consisting 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.

3.5 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.

3.6 Discussions

3.6.1 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-**3.3** in **Appendix**.

3.6.2 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[4], 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.

Another excluded example is ambiguity of cereal types[26], 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.

3.6.3 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[29], ARs (C19:0, C21:0, C23:0)[30]. However, according to current criterial of BFIRev[3] and validation[31], 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-3.1) 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.

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

- 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[30]. 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.

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[5]	DHBA DHPPA DHBA glycine 3,5-DHPPTA	HMDB0013677 HMDB0125533 InChi: QVGDKHUNWDVPOR-UHFFFAOYSA-N InChi: QHXNJIMVPAFCPR-UHFFFAOYSA-N	[29]
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	[30]

Table 3.2: "Geographically applicable (specific)" Biomarkers of Whole Grain Wheat Intake

3.6.4 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.

3.6.5 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.

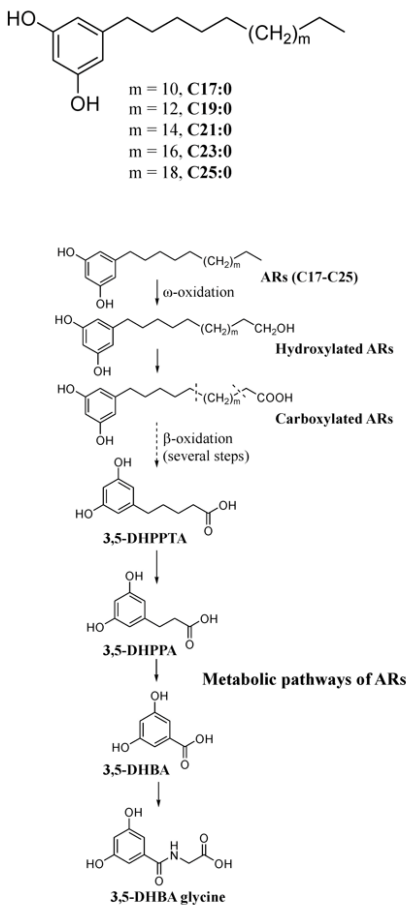

Figure 3.1: Structure of ARs and suggested metabolic pathway, adapted from [32]

3.7 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	[33]
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 3.3: Potential Biomarkers for WG Barley Intake

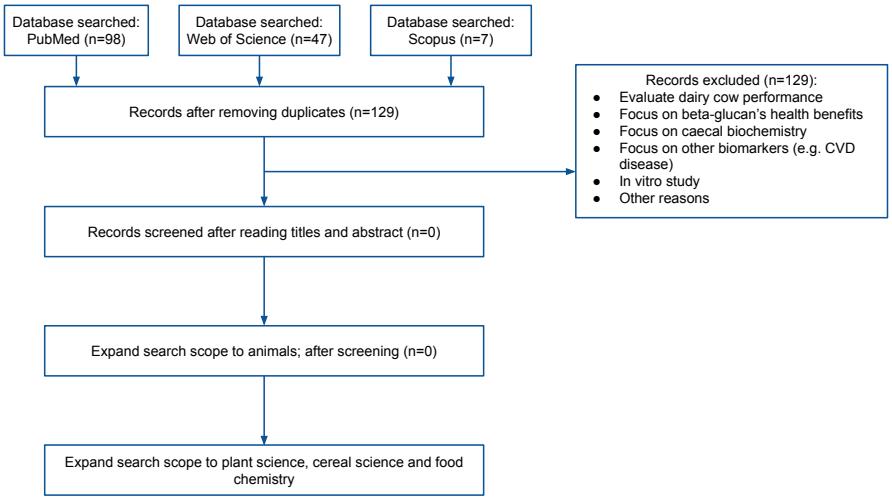

Figure 3.2: Flow Chart of Literature Searching and Screening for Articles of WG Barley Intake Biomarkers

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)	[34]
None-bread, White bread, WG bread	155	observation[6]	urine	HPLC-qTOF-MS	Benzoxazinoid-related metabolites (HHPAA, HBOA glycoside) ARs-related metabolites (DHPPA glucuronide, DHPPTA sulphate microbial-derived metabolites	[26]

Table 3.4: Reported markers distinguishing WG wheat intake, but NOT specific

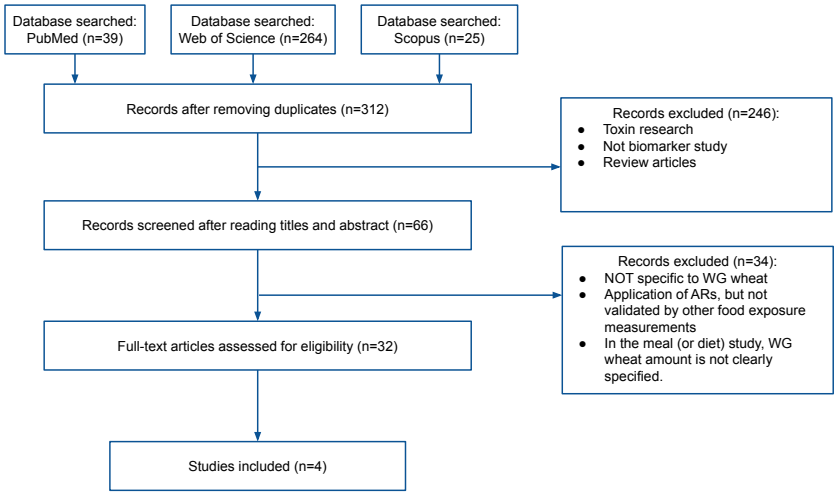

Figure 3.3: Flow Chart of Literature Searching and Screening for Articles of WG Wheat Intake Biomarkers

Chapter 4

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

Chapter 5

Urinary Intake Biomarker of Barley: Metabolite Identification

5.1 Abstract

5.2 Introduction

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

Phytosterols ubiquitously occur in plant-based food[35]. 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

5.2.2 β-glucuronidase for metabolites identification

Glucuronide group hurdles 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.

5.3 Materials and methods

5.3.1 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)

5.3.2 Apparatus

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

5.3.3 UPLC-MS/MS analysis of Sitostanol

5.3.4 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]
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

[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.

- Centrifuge at 3000 rmp for 3 min
- Add solvent A 300 µL
- Transfer supernatant to vials fur further LC-MS/MS analysis.

5.4 Results

5.4.1 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]-

5.4.2 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.

5.4.3

Chapter 6

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

6.1 Introduction

6.2 Materials and Methods

6.2.1 *In vitro* glucuronidation

[36]

6.2.2 Bioinformatics Tools and Software

Partition coefficient (logP and ClogP) was predicted by ChemDraw. MS^2 spectra was predicted by CFM-ID 3.0[37].

6.2.3 MS/MS

6.3 Results

6.3.1 Summary of Identification

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

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

6.3.2 Predicted logP value of AR metabolites

	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

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

6.3.3 *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-6.1).

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

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.

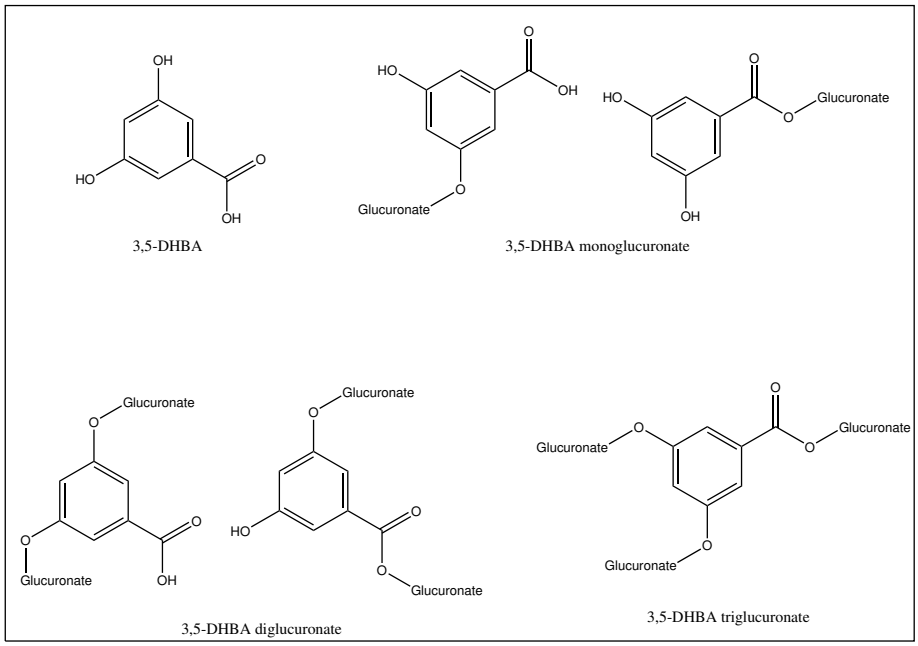

Figure 6.1: 3,5-DHBA

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 6.1: 3,5-DHBA monoglucuronate

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.

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

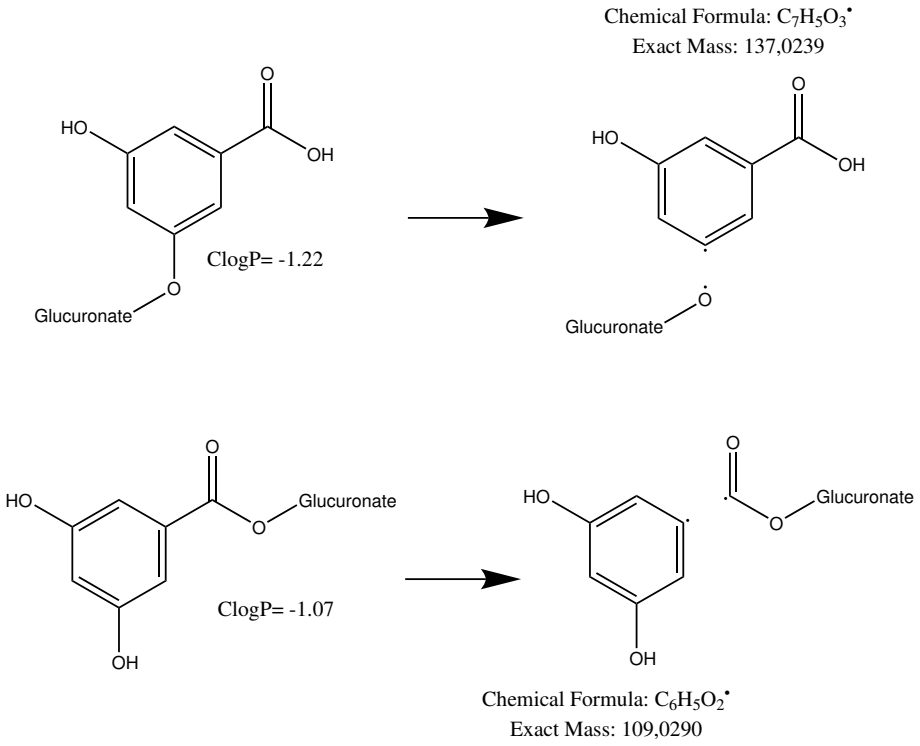

Figure 6.2: Fragmentation of 3,5-DHBA monoglucuronate

6.3.4 3,5-DHBA sulfate

6.3.5 Identifications of alkylresorcinol (AR) metabolites

6.4 Conclusion

6.5 Discussion

Chapter 7

Identification Strategies for BFIs Discovery

7.1 Introduction

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.

7.2 Materials and methods

In order to identify researches dealing with the identification of BFIs, we carried out an extensive literature search following the BFIRev methodology[3].

Briefly, searches were carried out in three databases (PubMed, Scopus, and ISI Web of Knowledge) in Jun 2019. In PubMed, the search terms were (nutrition*[Title/Abstract]) AND (biomarker*[Title] OR marker*[Title]) AND (validation*[Title/Abstract] OR validity*[Title/Abstract] OR validate*[Title/Abstract] OR assessment*[Title/Abstract]) NOT (animal OR rat OR mouse OR mice OR pig) NOT (disease*[Title] OR risk*[- Title] OR inflammat*[Title/Abstract] OR patient*[Title]). To avoid all the studies concerned with a single bio- marker while keeping studies on validation in general, we avoided using nutrient* or food* in the search strat- egy. The fields used for the other two databases were [Article Title/Abstract/Keywords] for Scopus and [Topic] for ISI Web of Science to replace [Title/Ab- stract] for PubMed. The search was limited to papers in English language and with no restriction applied for the publication dates. The review papers discussing the de- velopment and application of biomarkers in the nutri- tion field were selected in the process outlined in Fig. 1. The first draft scheme of validation criteria was based on criteria proposed in the review papers found by this literature search. This list was revised by three rounds of commenting by co-authors as well as feedback from pre- sentations at international conferences.

7.3 Discussion

7.3.1 Validation prior to Identification

In BFIs discovery pipeline, validation is the last step before a biomarker can be applied.

Howver, not all metabolites are identifiable in real world, for example, because the reference compound is not available.

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.

7.3.2 Challenges and Limitations

Open data

Bioinformatics tools

A lot of bioinformatics tools were emerged with the development of metabolomics. However, most bioinformatics tools were developed in data pre-processing, statistics,

Not enough bioinformatics tools were developed to facilitate the metabolite identification.

Chapter 8

Discussions

Chapter 9

Closing Remarks

9.1

Chapter 10

Appendix