LC-MS based Metabolomics: Biomarker Discover and Bioinformatics Tools

Tu Hu

Supervisor: Lars Ove Dragsted[1]
Co-supervisor: Gözde Gürdeniz[2]
Examiner: Ashfaq Ali[3]

Sep, 2019

[1]Professor, Department of Nutrition, Exercise and Sports, University of Copenhagen
[2]Senior researcher, COPSAC
[3]Bioinformatician, Department of Immunotechnology, University of Lund

Abstract

This calls for an abstract.

Contents

Chapter 1

A new chapter

Chapter 2

Introduction

2.1 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

2.2 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 3

Materials and Methods

3.1 Digital Literature Search and Systematic Review

3.2 Multi-variable Data Analysis

3.3 LC-MS Based Metabolome Profiling Systems

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

3.3.1 MSMS

3.4 title

Chapter 4

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

4.1 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 alkylresorcinols and homologous ratio of C17:0/C21:0 was suggested as a biomarker to measure whole grain wheat intake. Other promising intake biomarkers were summarized as well.

4.2 Introduction

Whole grains (WGs) contain a lot of non-nutrients in the bran. These non-nutrients 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].

However, 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[2]. Meanwhile, increasing evidence showed different WG types (such as wheat, rye, oat, barley

etc.) could benefit health differently.

Classical self-reported measurement tools (e.g. food diaries and food frequency questionnaires) used in observational studies could cause biases and confoundings in differentiating each cereal type.

Discovering BFIs of each whole grain type (e.g wheat, barley, rye, corn, rice etc.) 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 of the thesis work.

4.3 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. The keywords (animal* OR goat OR sheep OR cow OR mice OR mouse* OR animal model* OR dog*) were 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 presenting 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'.

4.4 Results

4.4.1 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(s)" 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 indicators of animal-source products[13, 14], which could also be regarded as *effect biomarkers* of animals. However, within above-mentioned results, intervention studies lacked objective markers for compliance monitoring.

4.4.2 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 observational 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.

Food items	No. subjects	Study design	Sample type	Analytical method	Candidate biomarker(s)	Identifier	Reference
WGs[1]	2845	Observational (11 countries)	fasting and non-fasting plasma	GC-MS	AR C17:0/C21:0	HMDB0038530 HMDB0031035	[15]
WG wheat WG rye	73	Observational (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	[16]
WG wheat WG rye	39	Intervention (cross-over)	plasma	GC-MS	Ratio of AR C17:0/C21:0	HMDB0038530 HMDB0031035	[17]
WG wheat WG rye	15	Intervention (cross-over)	plasma serum enterolactone	GC-MS	Ratio of AR C17:0/C21:0	HMDB0038530 HMDB0031035	[18]

Table 4.1: Biomarkers of Wheat Intake Reported in Intervention study

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 to measure 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 [19]. In grains, rye has homologous C17:0/C21:0 ratio close to 1.0, while wheat around 0.1, durum wheat around 0.01.

Further this marker was proposed by nutritionists to distinguish WG rye and wheat intake. In 2005, Linko[17] 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 validated this marker in healthy adults by an intervention study [18]. In plasma, the value was 0.1 after WG wheat intake, 0.6 after WG rye intake. In erythrocytes, the value was 0.06 and 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 Landberg[20]. But WG diet and refined diet can be distinguished by total ARs concentration in plasma.

EPIC[2] cohort study [15] further proved usefulness this marker in 2014. This observational 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

This biomarker showed its usefulness in type II diabetes study proving WG rye might better benefit type II diabetes prevention compared with WG wheat.

An observational study showed[21], in Chinese populations AR metabolite DHPPA was correlated with lower odds of type II diabetes and impaired glucose regulation. However, DHPPA as AR metabolite can originate from either whole grain wheat or rye. DHPPA can only indicate total intake of WG rye and wheat. But, this marker cannot provide details of each whole grain type intake.

Other two researches better showed 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 implied that, a whole grain diet dominated by rye could be favourable for type II diabetes prevention.

In EPIC cohort study[15], there was a very interesting phenomena. Rye has higher constitutions in Danish populations' WG source (70%) than Swedish (55%) on average. However, Danish participants in EPIC cohort showed lower AR C17:0/C21:0 ratio (0.37) than Swedish participants (0.43). Swedish participants in EPIC cohort studies were healthy adults, while Danish participants were obese or over-weights subjects. Those participants may have different dietary habits and consume less than average wg rye. This may also imply rye could be favourable in weight control.

[2]European Prospective Investigation into Cancer and Nutrition

These studies indicated that, application of biomarkers for each type of cereal intake could provide more information to study each cereal type'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* such as microbioal derivitives[26]. These compounds do not exclusively present in WG wheat. Therefore, they can not specifically indicate WG wheat intake. These results were summarized in Appendix.

However, another study[28] proposed using a panel of metabolites consisting 7 AR metabolites, 5 BXO metabolites and 5 phenolic acid derivatives to objectively assess WG wheat intake. Because concentration of these phytochemicals vary in different cereal grains. Therefore, a combination of their metabolites could potentially indicate intake of different cereals. However, this conclusion needs to be further validated in other intervention and observation studies.

4.5 Conclusions

None biomarker has been reported for WG barley intake both from human and animal studies.

Total ARs from plasma can indicate WG rye and wheat intake. The homologues ratio of ARs C17:0/C21:0 can indicate relative composition of WG rye and wheat. Combing these two markers can objectively indicate WG wheat intake. This marker has been implemented in several studies and proven its usefulness.

4.6 Discussions

4.6.1 Potential BFIs for WG barley

Search results in food chemistry, cereal science and plant science database showed some compounds presenting exclusively in WG barley. These could give hints for further identification. The results were summarized in Table-**3.3** in **Appendix**.

4.6.2 Exclusion Criteria of WG Wheat intake biomarker

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

We used a strict exclusion criteria in order to identify specific biomarkers to distinguish WG wheat intake from refined wheat or other WG types:

1. The terms “WGs or dietary fibers or cereal fibers or bread” were used. However the details about how much WG wheat wasn't included.
2. In observational studies, food intake was only measured by biomarkers without any validations (such as FFQ, FR etc)
3. Interventions were WG diet or meal (including several WG types). Biomarkers were not specific to wheat.

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.

4.6.3 AR Metabolites for WG Wheat Intake and Geographical Applicabilities

Biomarkers for WG wheat intake could have their geographical applicabilities. In this context, “geography” refers to a region 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, 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, DHPPTA: They are phase I metabolites of AR(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.
- 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 intake can also produce these metabolites.

To our best knowledges, rye is rarely consumed in USdue to its limited availabilities. Therefore, in such a “geography” where WG wheat is the only AR source. These markers have their geographical applicabilities though they might not be applied in both WG rye and wheat consuming areas.

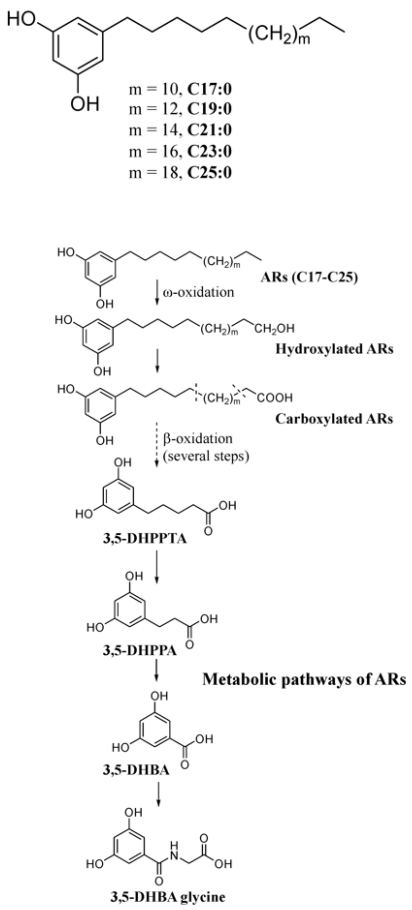

Figure 4.1: Structure of ARs and suggested metabolic pathway, adapted from [32]

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[3]	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 4.2: "Geographically applicable (specific)" Biomarkers of Whole Grain Wheat Intake

4.6.4 Ambiguous Use of the Term "Biomarker"

Ambiguously using the term "biomarker" in the publications hurdled scientific communications. Most retrieved results mentioned "biomarker". However, "biomarkers" referred different concepts in different contexts, e.g. *food intake biomarkers*, *food compound intake biomarkers* and *effect biomarkers*. In literature search phase of this review, it is difficult for us to quickly classify biomarkers before reading the full text. Correctly using these terms might reduce confusions and make communications easier.

Gao[9] and Dragsted[8] proposed the new ontology and classification schema of "Biomarker". The awareness and implementations of the new ontology and classification schema will relief this problem.

4.7 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 4.3: Potential Biomarkers for WG Barley Intake

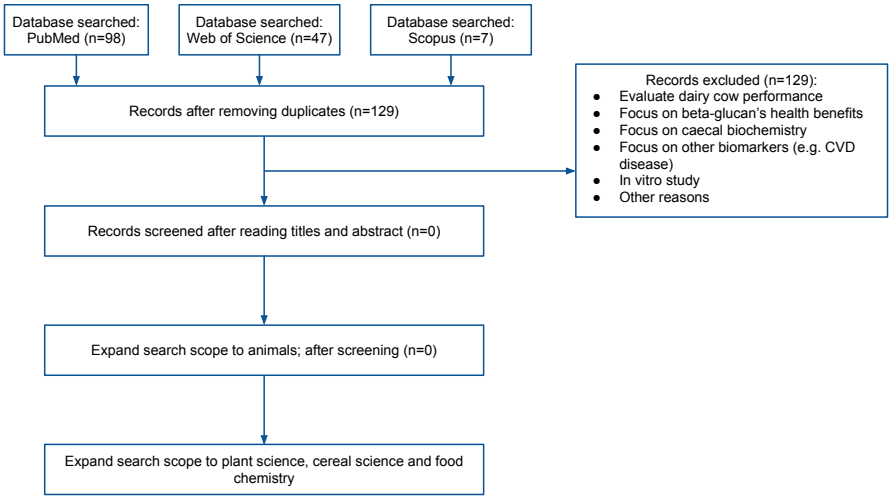

Figure 4.2: Flow Chart of Literature Searching and Screening for Articles of WG Barley Intake Biomarkers

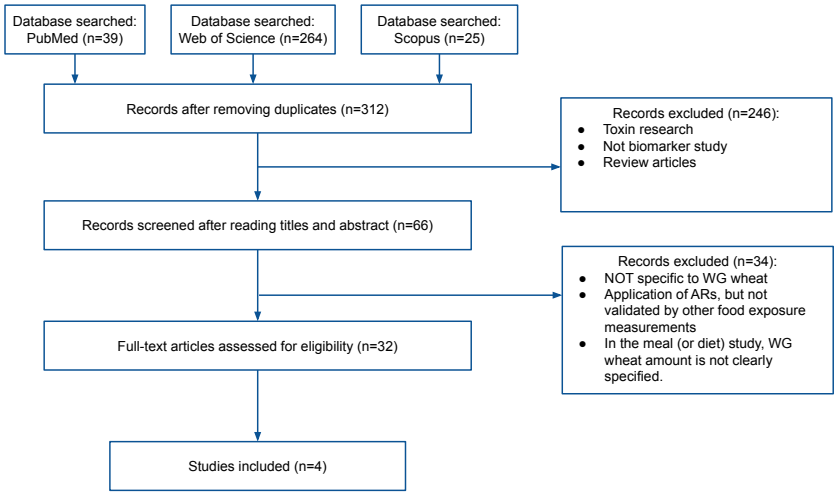

Figure 4.3: Flow Chart of Literature Searching and Screening for Articles of WG Wheat Intake Biomarkers

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)	[34]
None-bread, White bread, WG bread	155	observation[4]	urine	HPLC-qTOF-MS	Benzoxazinoid-related metabolites (HHPAA, HBOA glycoside) ARs-related metabolites (DHPPA glucuronide, DHPPTA sulphate microbial-derived metabolites	[26]
WGs[5]	266	randomized, parallel-group, intervention	plasma	GC-MS	Total ARs	[35]

Table 4.4: Reported markers distinguishing WG wheat intake, but NOT specific

Chapter 5

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

Chapter 6

Urinary Intake Biomarker of Barley: Metabolite Identification

6.1 Abstract

6.2 Introduction

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

Phytosterols ubiquitously occur in plant-based food[36]. 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

6.2.2 β-glucuronidase for metabolites identification

β-glucuronidase catalyses β-glucuronates hydrolysis. This enzyme is routinely used for enzymatic hydrolysis of urine, plasma and other fluids prior to analysis by enzyme immunoassay or mass spectrometry etc.

β-glucuronates are common phase II metabolites presenting in urine and plasma.

6.3 Materials and methods

6.3.1 Chemicals

Sitostanol reference compound (CAS: 83-45-4, Avanti Polar Lipids Inc., USA), β-glucuronidase (CAS: 9001-45-0, E.C. number 3.2.1.31, Sigma-Aldrich, from *Escherichia Coil*)

6.3.2 Apparatus

6.3.3 β-glucuronidase experiment

Urine samples were treated with β-glucuronidase for 1.5 h to hydrolyse glucuronate group. A positive control was used. Details were described in **Appendix**.

6.3.4 UPLC-MS/MS analysis of Sitostanol

6.4 Results

6.4.1 Retention Time (RT) and m/z in Different Matrix of Sitostanol

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]-

6.4.2 β-glucuronidase Experiment

β-glucuronidase hydrolysed both ions. The glucuronates decreased intensities.

The expected unglucuronated ion (m/z 325.2739) was not detected. the reason could be: (1) not ionized.

Another expected unglucuronated ion (m/z 341.2675) was detected on the same day of experiment in full scan mode. However, this ion was not detected after one week storage in -20 freezer. This could be the reason of degradation. Ion 341.2675 was detected in RT 0.88 indicating it is a highly polar compound. Its structure needs to be further confirmed by MS/MS analysis.

6.5 Discussion

Chapter 7

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

7.1 Introduction

7.2 Materials and Methods

7.2.1 Analysis of *in vitro* glucuronidation samples

Gözde conducted *In vitro* glucuronidation experiments. Detailed methods were described in Appendix.
[37]

7.2.2 Bioinformatics Tools and Software

Partition coefficient (logP and ClogP) was predicted by ChemDraw. MS^2 spectra was predicted by CFM-ID 3.0[38].

7.2.3 MS/MS

7.3 Results

7.3.1 Summary of Identification

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

No.	m/z	RT (Quat)	RT (Bi)	MS/MS	Annotation	Suggested Compound	Level of Identification
	329.0582	2.54	0.97	175.02 153.02 113.02 96.96	$[M-H]^-$	3,5-DHBA glucuronate	I
	329.0582	3.51	1.12		$[M-H]^-$	3,5-DHBA glucuronate	I
	232.9750	3.58	1.37	215.09 153.01 109.03 96.96	$[M-H]^-$	3,5-DHBA sulfate	II
	210.0411	4.64	1.48		$[M-H]^-$	3,5-DHBA glycine	II
	153.0187	1.18	1.88	109.03	$[M-H]^-$	3,5-DHBA	II
	357.0898	4.24	1.91		$[M-H]^-$	3,5-DHPPA glucuronate	
	261.0077	4.32	2.10		$[2M-H]^-$	3,5-DHPPA sulfate	
	263.022	5.48	3.31	247.99 183.06 168.04 153.02 79.96	$[M-H]^-$	Unknown sulfate	

7.3.2 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.

7.3.3 *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-DHBA + GluA - H]$^-$ (glucuronated at carboxylic acid hydroxyl group)
3.97	329.05	[3,5-DHBA + GluA - H]$^-$ (glucuronated at benzene hydroxyl group)
3.6	252.04	[3,5-DHBA + 2GluA - 2H]$^{2-}$
7.34	339.55	[3,5-DHBA + 3GluA - 2H]$^{2-}$

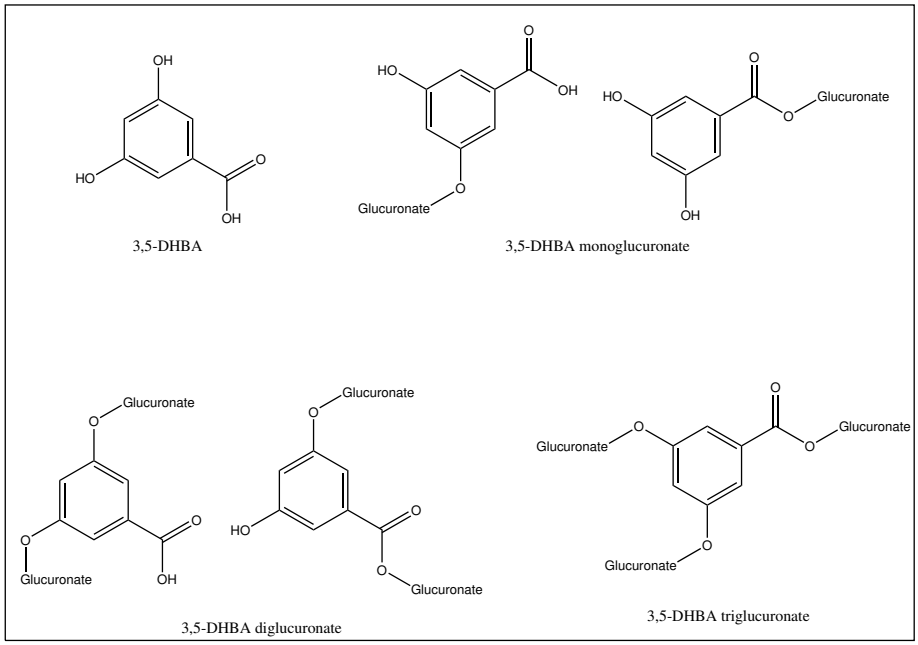

Figure 7.1: 3,5-DHBA

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.

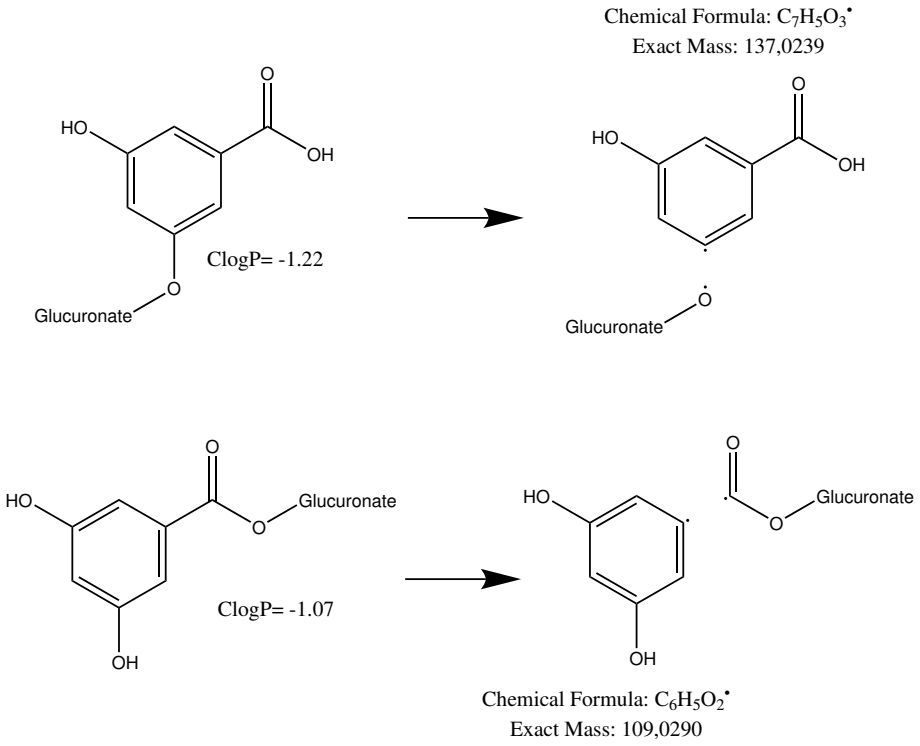

Figure 7.2: Fragmentation of 3,5-DHBA monoglucuronate

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 7.1: 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-.

7.3.4 3,5-DHBA sulfate

7.4 Conclusion

7.5 Discussion

Chapter 8

Identification Strategies for BFIs Discovery

8.1 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.

8.2 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.

8.3 Discussion

8.3.1 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.

8.3.2 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 9

Automate Identification based on KUDB (Københavns University Data Base)

Chapter 10

Discussions

Chapter 11

Closing Remarks

11.1

Chapter 12

Appendix

12.1 β-glucuronidase Experiment

12.1.1 Experimental Procedures

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

[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.

12.1.2 Chromatograms

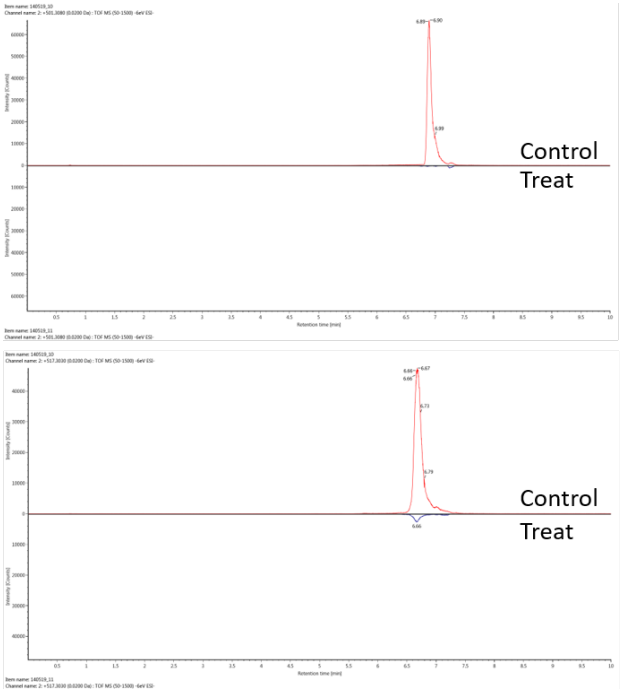

Figure 12.1: Chromatograms of glucuronates (control and treat)

Bibliography

1. Piepoli, M. F. *et al.* 2016 European Guidelines on cardiovascular disease prevention in clinical practice. *Atherosclerosis* **252,** 207–274. ISSN: 00219150. http://www.ncbi.nlm.nih.gov/pubmed/27664503%20https://linkinghub.elsevier.com/retrieve/pii/S0021915016302143 (Sept. 2016).
2. Aune, D. *et al.* Whole grain consumption and risk of cardiovascular disease, cancer, and all cause and cause specific mortality: systematic review and dose-response meta-analysis of prospective studies. *BMJ (Clinical research ed.)* **353,** i2716. ISSN: 1756-1833. http://www.ncbi.nlm.nih.gov/pubmed/27301975%20http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4908315 (June 2016).
3. Praticò, G. *et al.* Guidelines for Biomarker of Food Intake Reviews (BFIRev): How to conduct an extensive literature search for biomarker of food intake discovery. *Genes and Nutrition* **13.** ISSN: 18653499 (2018).
4. Wishart, D. S. *et al.* HMDB 4.0: the human metabolome database for 2018. eng. *Nucleic acids research* **46,** D608–D617. ISSN: 1362-4962 (Electronic) (Jan. 2018).
5. *FooDB* http://foodb.ca/.
6. Medina-Remón, A. *et al.* Phenol-Explorer 3.0: a major update of the Phenol-Explorer database to incorporate data on the effects of food processing on polyphenol content. *Database* **2013.** ISSN: 1758-0463. https://dx.doi.org/10.1093/database/bat070 (2013).
7. Dictionary of food compounds with CD-ROM. *Choice Reviews Online.* ISSN: 0009-4978 (2013).
8. Dragsted, L. O. *et al.* Dietary and health biomarkers—time for an update. *Genes and Nutrition* **12,** 1–7. ISSN: 18653499 (2017).
9. Gao, Q. *et al.* A scheme for a flexible classification of dietary and health biomarkers. *Genes & Nutrition* **12,** 34. ISSN: 1865-3499. https://doi.org/10.1186/s12263-017-0587-x (Dec. 2017).

10. Bird, A. R. *et al.* Wholegrain foods made from a novel high-amylose barley variety (Himalaya 292) improve indices of bowel health in human subjects. eng. *The British journal of nutrition* **99,** 1032–1040. ISSN: 0007-1145 (Print) (May 2008).

11. Ames, N. *et al.* A double-blind randomised controlled trial testing the effect of a barley product containing varying amounts and types of fibre on the postprandial glucose response of healthy volunteers. eng. *The British journal of nutrition* **113,** 1373–1383. ISSN: 1475-2662 (Electronic) (May 2015).

12. Marungruang, N., Tovar, J., Bjorck, I. & Hallenius, F. F. Improvement in cardiometabolic risk markers following a multifunctional diet is associated with gut microbial taxa in healthy overweight and obese subjects. eng. *European journal of nutrition* **57,** 2927–2936. ISSN: 1436-6215 (Electronic) (Dec. 2018).

13. Abidi, S., Ben Salem, H., Vasta, V. & Priolo, A. Supplementation with barley or spineless cactus (Opuntia ficus indica f. inermis) cladodes on digestion, growth and intramuscular fatty acid composition in sheep and goats receiving oaten hay. *SMALL RUMINANT RESEARCH* **87,** 9–16. ISSN: 0921-4488 (Nov. 2009).

14. Foster, A. P. *et al.* Serum IgE and IgG responses to food antigens in normal and atopic dogs, and dogs with gastrointestinal disease. eng. *Veterinary immunology and immunopathology* **92,** 113–124. ISSN: 0165-2427 (Print) (May 2003).

15. Kyro, C. *et al.* Plasma alkylresorcinol concentrations, biomarkers of wholegrain wheat and rye intake, in the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort. *BRITISH JOURNAL OF NUTRITION* **111,** 1881–1890. ISSN: 0007-1145 (May 2014).

16. Landberg, R., Aman, P., Hallmans, G. & Johansson, I. Long-term reproducibility of plasma alkylresorcinols as biomarkers of whole-grain wheat and rye intake within Northern Sweden Health and Disease Study Cohort. *EUROPEAN JOURNAL OF CLINICAL NUTRITION* **67,** 259–263. ISSN: 0954-3007 (Mar. 2013).

17. Raninen, K. J. *et al.* Fiber content of diet affects exhaled breath volatiles in fasting and postprandial state in a pilot crossover study. *NUTRITION RESEARCH* **36,** 612–619. ISSN: 0271-5317 (June 2016).

18. Adlercreutz, H., Peñalvo, J. L., Linko-Parvinen, A.-M., Tikkanen, M. J. & Landberg, R. Alkylresorcinols from Whole-Grain Wheat and Rye Are Transported in Human Plasma Lipoproteins. *The Journal of Nutrition* **137,** 1137–1142. ISSN: 0022-3166. `https://dx.doi.org/10.1093/jn/137.5.1137` (2007).

19. Chen, Y., Ross, A. B., Åman, P. & Kamal-Eldin, A. Alkylresorcinols as Markers of Whole Grain Wheat and Rye in Cereal Products. *Journal of Agricultural and Food Chemistry* **52,** 8242–8246. ISSN: 0021-8561. `https://doi.org/10.1021/jf049726v` (Dec. 2004).

20. Landberg, R., Kamal-Eldin, A., Andersson, A., Vessby, B. & Aman, P. Alkylresorcinols as biomarkers of whole-grain wheat and rye intake: plasma concentration and intake estimated from dietary records. *AMERICAN JOURNAL OF CLINICAL NUTRITION* **87,** 832–838. ISSN: 0002-9165 (Apr. 2008).

21. Sun, T. *et al.* Plasma Alkylresorcinol Metabolite, a Biomarker of Whole-Grain Wheat and Rye Intake, and Risk of Type 2 Diabetes and Impaired Glucose Regulation in a Chinese Population. *DIABETES CARE* **41,** 440–445. ISSN: 0149-5992 (Mar. 2018).

22. Magnusdottir, O. K. *et al.* Plasma alkylresorcinols C17:0/C21:0 ratio, a biomarker of relative whole-grain rye intake, is associated to insulin sensitivity: a randomized study. *EUROPEAN JOURNAL OF CLINICAL NUTRITION* **68,** 453–458. ISSN: 0954-3007 (Apr. 2014).

23. Biskup, I. *et al.* Plasma alkylresorcinols, biomarkers of whole-grain wheat and rye intake, and risk of type 2 diabetes in Scandinavian men and women. *AMERICAN JOURNAL OF CLINICAL NUTRITION* **104,** 88–96. ISSN: 0002-9165 (July 2016).

24. Bresciani, L. *et al.* Bioavailability and metabolism of phenolic compounds from wholegrain wheat and aleurone-rich wheat bread. *MOLECULAR NUTRITION & FOOD RESEARCH* **60,** 2343–2354. ISSN: 1613-4125 (Nov. 2016).

25. Jensen, B. M. *et al.* Quantitative analysis of absorption, metabolism, and excretion of benzoxazinoids in humans after the consumption of high- and low-benzoxazinoid diets with similar contents of cereal dietary fibres: a crossover study. *EUROPEAN JOURNAL OF NUTRITION* **56,** 387–397. ISSN: 1436-6207 (Feb. 2017).

26. Garcia-Aloy, M. *et al.* Nutrimetabolomics fingerprinting to identify biomarkers of bread exposure in a free-living population from the PREDIMED study cohort. *METABOLOMICS* **11,** 155–165. ISSN: 1573-3882 (Feb. 2015).

27. Nybacka, S., Forslund, H. B. & Hedelin, M. Validity of a web-based dietary questionnaire designed especially to measure the intake of phytoestrogens. *JOURNAL OF NUTRITIONAL SCIENCE* **5.** ISSN: 2048-6790 (Sept. 2016).

28. Zhu, Y., Wang, P., Sha, W. & Sang, S. Urinary Biomarkers of Whole Grain Wheat Intake Identified by Non-targeted and Targeted Metabolomics Approaches. *SCIENTIFIC REPORTS* **6.** ISSN: 2045-2322 (Nov. 2016).

29. Zhu, Y., Shurlknight, K. L., Chen, X. & Sang, S. Identification and Pharmacokinetics of Novel Alkylresorcinol Metabolites in Human Urine, New Candidate Biomarkers for Whole-Grain Wheat and Rye Intake. *JOURNAL OF NUTRITION* **144,** 114–122. ISSN: 0022-3166 (Feb. 2014).

30. McKeown, N. M. *et al.* Comparison of plasma alkylresorcinols (AR) and urinary AR metabolites as biomarkers of compliance in a short-term, whole-grain intervention study. *EUROPEAN JOURNAL OF NUTRITION* **55,** 1235–1244. ISSN: 1436-6207 (Apr. 2016).

31. Dragsted, L. O. *et al.* Validation of biomarkers of food intake—critical assessment of candidate biomarkers. *Genes & Nutrition* **13,** 14. ISSN: 1865-3499. https://doi.org/10.1186/s12263-018-0603-9 (2018).

32. Sang, S. Biomarkers of Whole Grain Intake. *JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY* **66,** 10347–10352. ISSN: 0021-8561 (Oct. 2018).

33. Gurdeniz, G. *et al.* Detecting Beer Intake by Unique Metabolite Patterns. eng. *Journal of proteome research* **15,** 4544–4556. ISSN: 1535-3907 (Electronic) (Dec. 2016).

34. Keaveney, E. M. *et al.* Postprandial plasma betaine and other methyl donor-related responses after consumption of minimally processed wheat bran or wheat aleurone, or wheat aleurone incorporated into bread. *BRITISH JOURNAL OF NUTRITION* **113,** 445–453. ISSN: 0007-1145 (Feb. 2015).

35. Ross, A. B. *et al.* Plasma alkylresorcinols as a biomarker of whole-grain food consumption in a large population: results from the WHOLEheart Intervention Study. *AMERICAN JOURNAL OF CLINICAL NUTRITION* **95,** 204–211. ISSN: 0002-9165 (Jan. 2012).

36. Muenger, L. H., Boulos, S. & Nystroem, L. UPLC-MS/MS Based Identification of Dietary Steryl Glucosides by Investigation of Corresponding Free Sterols. *FRONTIERS IN CHEMISTRY* **6.** ISSN: 2296-2646 (Aug. 2018).

37. Liu, Z. & Franklin, M. R. Separation of four glucuronides in a single sample by high-pressure liquid chromatography and its use in the determination of UDP glucuronosyltransferase activity toward four aglycones. *Analytical Biochemistry* **142,** 340–346. ISSN: 0003-2697. https://www.sciencedirect.com/science/article/pii/0003269784904743 (Nov. 1984).

38. Djoumbou-Feunang, Y. *et al.* CFM-ID 3.0: Significantly Improved ESI-MS/MS Prediction and Compound Identification. *Metabolites* **9.** ISSN: 2218-1989. https://www.mdpi.com/2218-1989/9/4/72 (2019).