Title

Characterization of Food Carbohydrates by Liquid Chromatography-Mass Spectrometry Methods

Permalink

https://escholarship.org/uc/item/8cs3g77c

Author

Castillo, Juan Jose

Publication Date

2021

Peer reviewed|Thesis/dissertation

Characterization of Food Carbohydrates by Liquid Chromatography-Mass Spectrometry Methods

By

JUAN JOSE CASTILLO

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Chemistry

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

_____

Carlito B. Lebrilla, Chair

_____

Donald P. Land

_____

Susan E. Ebeler

Committee in Charge

2022

DEDICATION

I dedicate this research thesis to my mother, Rosa Alvarez Castañeda (descanse en paz), my

father, Pedro Castillo Herrera, my sister, Rohana Castillo Alvarez, my brother, Manuel Castillo

Alvarez, and family members. Thank you for all your care, love, and support.

Juan Jose Castillo Alvarez

August 2021

Davis, California

ACKNOWLEDGEMENTS

I would first like to thank my parents for leaving their home in Mexico and moving to the United States of America for a better life. I would like to thank my mother (rest in peace mama) for single handedly raising me and my siblings in our studio apartment. I would like to thank my father for all always being there and providing love and support. My siblings, Rohana and Manuel Castillo for their unconditional love, support, and encouragement. I would like to thank my sister-in-law Virginia Mira to all the love and support. I would like to thank Megan L. Tesene for all the love, support, and being there during difficult times. I would like to thank the neighborhood I grew in, "The Canal," for experiencing life from a hardship perspective. I like to acknowledge my neighbors and family members who have helped me throughout my childhood.

I would like to acknowledge all my teachers and mentors. I would like to thank Sue Herman, my high school teacher, who always believed in me, Dr. Jennifer Loeser my general chemistry community college professor, Dr. Scott Serafin my organic chemistry professor who first introduced me to mass spectrometry, Dr. Peter Palmer who encouraged me to join his research lab at SFSU and to pursue a graduate degree, Dr. Ten-Yang Yen who ran LC-MS experiments with me at SFSU, Dr. Leticia Márquez-Magaña for the opportunity to join the health equity research (HER) lab, the HER lab members (Dr. Cathy Samayoa, Rebecca, Saba, and Edgar) for the help and support, SF BUILD for the encouragement to take on graduate school, Dr. Urmimala Sarkar for the summer research opportunity at UCSF, and Dr. Kala Mehta who helped network and apply to graduate school.

I would like to thank my research advisor, Dr. Carlito B. Lebrilla for all the support and the opportunity to join his lab. Thank you for all the training, mentorship, project opportunities, and for building all those memories. During my time as graduate student, I have learned so much about research and many items beyond the lab. It truly was a dream of mine to be in Davis and work in his lab.

I would like to thank my previous colleagues Dr. Ace Galermo, Dr. Matthew Amicucci, and Dr. Eshani Nandita who mentored me in the beginning of my graduate school time. I want to thank my current colleagues whom I closely worked with and grew as a scientist in many ways, Garret Couture, and Nikita Bacalzo. I would like to thank Ye (Winnie) Chen for her amazing support on all the project presented in this dissertation. I would like to thank my undergrads Lorraine Ishida, and Matthew Sujanto for all the help. Lastly, many thanks to whole Lebrilla lab for all the support Dr. Axe Xie, Jennyfer Tena, Qing Zhou, Ying Sheng, Christopher Ranque, Siyu Chen, Anita Vinjamuri, Armin Oloumi, Ryan Schindler, Aaron Stacy, Yasmine Bouchibti, and Cheng-Yu (Charlie) Weng.

I would like to thank my dissertation committee members Dr. Susan Ebeler and Dr. Donald Land for their support on my qualifying exam and dissertation.

Lastly, I would like to thank member from the Department of Chemistry, Scott Berg, Brad Wolf (previous member), and Albert Sy for their generous help.

Characterization of Food Carbohydrates by

Liquid Chromatography-Mass Spectrometry Methods

ABSTRACT

Carbohydrates are the most abundant biomolecules in nature, and they are involved in many biological functions. These biomolecules play a key role in shaping the human gut microbiome. However, little is known about how the specific structures of food carbohydrates mediate this link between the gut microbiome and health. Characterizing carbohydrate structures remains a significant challenge, and their compositions in food remain unknown. The chapters presented in this dissertation include a series of advance liquid chromatography-mass spectrometry (LC-MS) methods to characterize food carbohydrates. Chapter I includes structures, functions, and biosynthesis of plant carbohydrates. In addition, the chapter describes the functions of dietary carbohydrates serving as prebiotics followed by the modern LC-MS methods to characterize carbohydrates at the monosaccharide, glycosidic linkage, and sequence levels. Chapter II presents a novel high-throughput ultra-high performance liquid chromatography triple quadrupole mass spectrometry (UHPLC-QqQ MS) workflow to determine the absolute quantitation of monosaccharide composition in 850 foods. The resulting monosaccharide compositions were used to generate a glycan encyclopedia (Glycopedia) for feeding studies and personalizing diets. Chapter III presents the development of a *de novo* multidimensional LC-MS workflow to structurally elucidate oligosaccharides derived from plant polysaccharides at the sequence, monosaccharide, and glycosidic linkage levels. Chapter IV presents a glycosidic linkage method to elucidate linkages in food and feces for monitoring carbohydrate-microbe interactions in clinical feeding studies involving mice and humans.

PUBLICATIONS

1.   Castillo, J. J.; Galermo, A. G.; Amicucci, M. J.; Nandita, E.; Couture, G.; Bacalzo, N.; Chen, Y.; Lebrilla, C. B., A Multidimensional Mass Spectrometry-Based Workflow for De Novo Structural Elucidation of Oligosaccharides from Polysaccharides. *Journal of the American Society for Mass Spectrometry* **2021**, *32* (8), 2175-2185.

2.   Delannoy-Bruno, O.; Desai, C.; Raman, A. S.; Chen, R. Y.; Hibberd, M. C.; Cheng, J.; Han, N.; Castillo, J. J.; Couture, G.; Lebrilla, C. B., Evaluating microbiome-directed fibre snacks in gnotobiotic mice and humans. *Nature* **2021**, 1-5.

3.   Patnode, M. L.; Guruge, J. L.; Castillo, J. J.; Couture, G. A.; Lombard, V.; Terrapon, N.; Henrissat, B.; Lebrilla, C. B.; Gordon, J. I., Strain-level functional variation in the human gut microbiota based on bacterial binding to artificial food particles. *Cell Host & Microbe* **2021**, *29* (4), 664-673. e5.

4.   Amicucci, M. J.; Nandita, E.; Galermo, A. G.; Castillo, J. J.; Chen, S.; Park, D.; Smilowitz, J. T.; German, J. B.; Mills, D. A.; Lebrilla, C. B., A nonenzymatic method for cleaving polysaccharides to yield oligosaccharides for structural analysis. *Nature Communications* **2020**, *11* (1), 1-12.

5.   Galermo, A. G.; Nandita, E.; Castillo, J. J.; Amicucci, M. J.; Lebrilla, C. B., Development of an extensive linkage library for characterization of carbohydrates. *Analytical Chemistry* **2019**, *91* (20), 13022-13031.

ABBREVIATIONS

ADP: Adenosine diphosphate

All: Allose

Ara: Arabinose

CAZymes: Carbohydrate-active enzymes

CID: Collision induced dissociation

dMRM: Dynamic multiple reaction monitoring

DoPe: Degree of polymerization

DMSO-Dimethyl Sulfoxide

DP: Degree of polymerization

EIC: Extracted ion chromatogram

FOS: Fructooligosaccharides

Fru: Fructose

Fuc: Fucose

Gal: Galactose

GalA: Galacturonic acid

GalNAc: *N*-actylgalactosamine

GAX: Glucoronoaraninoxylan

GC-MS: Gas chromatography-mass spectrometry

GDP: Guanosine diphosphate

GH: Glycosyl hydrolyses

Glc: Glucose

GlcA: Glucuronic acid

GlcNAc: *N*-actylgalactosamine

GOS: Galactooligosaccharides

GTs: Glycosyltransferases

HG: Homogalacturonan

HFCS: High fructose corn syrup

HILIC: Hydrophilic interaction liquid chromatography

HPLC: High-pressure liquid chromatography

LC: Liquid chromatography

LC-MS: Liquid chromatography-mass spectrometry

LC-MS/MS: Liquid chromatography-tandem mass spectrometry

Man: Mannose

MRM: Multiple reaction monitoring

MS: Mass spectrometry

MS/MS: Tandem mass spectrometry

NaOH: Sodium hydroxide

PGC: Porous graphitized carbon

PL: Polysaccharide lyases

PMP: 1-phenyl-3-methyl-5-pyrazolone

QqQ-MS: Triple quadrupole mass spectrometry

Q-TOF-MS: Quadrupole time-of-flight mass spectrometry

RFO: Raffinose family oligosaccharides

RGI: Rhamnogalacturonan I

RGII: Rhamnogalacturonan II

Rha: Rhamnose

Rib: Ribose

SCFA: Short chain-fatty acids

SPE: Solid-phase extraction

SRM: Selected reaction monitoring

TFA: Trifluoroacetic acid

UDP: Uracil-diphosphate

UHPLC-QqQ MS: Ultra-high pressure liquid chromatography/triple quadrupole mass
Spectrometry

XG: Xyloglucan

XOS: Xylooligosaccharides

Xyl: Xylose

TALBE OF CONTENTS

Chapter 1

Introduction to Food Carbohydrate Analysis by

Liquid Chromatography-Mass Spectrometry

Structure, Functions, and Biosynthesis of Plant Carbohydrates

Structure

Food carbohydrates are a class of macromolecules composed of carbon, hydrogen, oxygen, and at times containing nitrogen. These biomolecules possess hydrogens and hydroxyl groups attached to the carbon backbone to form ring structures in pyranose (6-membered) or furanose (5-membered) forms.[1] The building blocks for carbohydrates in their simplest form are monosaccharides.[2] Chemically, monosaccharides with aldehydes at their reducing ends are aldose saccharides, while monosaccharides that contain ketones are ketose saccharides. Common monosaccharides found in plants belong to one of several classes of: hexoses, pentoses, deoxyhexoses, hexuronic acids, and *N*-acetylhexosamines (**Figure 1.1**). The monosaccharides found in plants include glucose (Glc), galactose (Gal), fructose (Fru), mannose (Man), ribose (Rib), xylose (Xyl), arabinose (Ara), fucose (Fuc), rhamnose (Rha), *N*-acetylglucosamine (GlcNAc), *N*-acetylgalactosamine (GalNAc), glucuronic acid (GlcA), and galacturonic acid (GalA). Common monosaccharide names, abbreviations, cartoon representation, and structures found in plant carbohydrates are listed in **Table 1.1**.
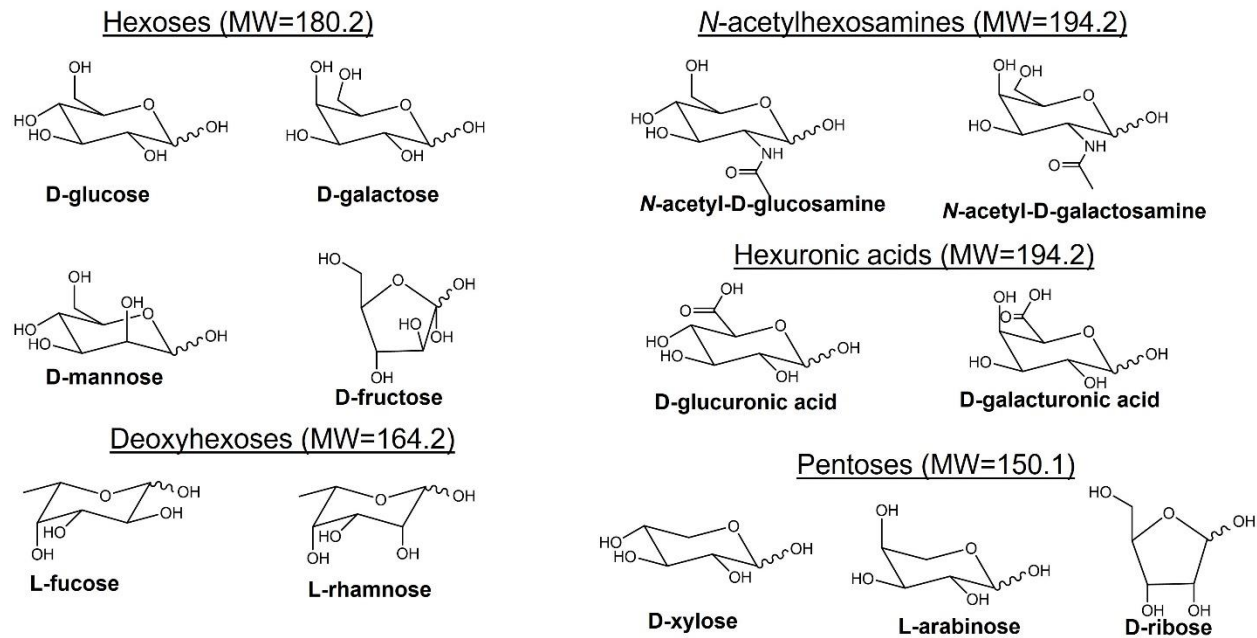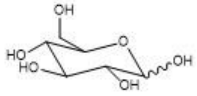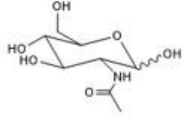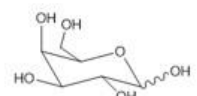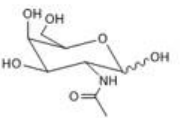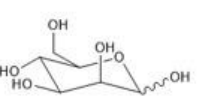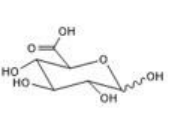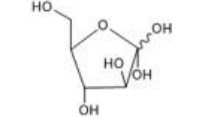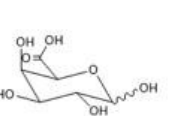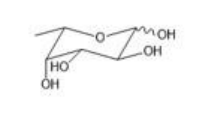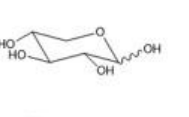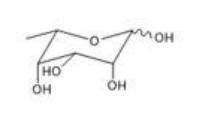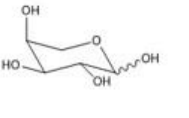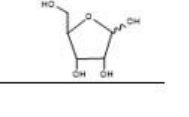
Figure 1.1. Monosaccharide structures that are commonly found in plant-based foods.

Table 1.1 Common plant monosaccharides with their corresponding names, abbreviations, cartoon symbols, and structures.

name	abbreviation	symbol	structure	name	abbreviation	symbol	structure
D-glucose	Glc			N-acetyl-D-glucosamine	GlcNAc		
D-galactose	Gal			N-acetyl-D-galactosamine	GalNAc		
D-mannose	Man			D-glucuronic acid	GlcA		
D-fructose	Fru			D-galacturonic acid	GalA		
L-fucose	Fuc			D-xylose	Xyl		
L-rhamnose	Rha			L-arabinose	Ara		
				D-ribose	Rib		

Carbohydrates in food exhibit a broad range of degrees of polymerization including monosaccharides, disaccharides, oligosaccharides, and polysaccharides.[3] In the human diet, carbohydrates account for anywhere between 40-80% of the energy needed to fuel our bodies.[4, 5] In the food industry, large molecular weight carbohydrates provide unique textures, sweetness, and viscosity to processed foods while providing many functions.[6, 7] For example, xanthan gum is a polysaccharide commonly used as a food additive serving as an effective thickening agent

and stabilizer to prevent ingredients from separating.[8] On the other hand, products containing smaller molecular weight carbohydrates such as high fructose corn syrup (HFCS) are added to food as a low cost alternative for adding sweetness to products.[9] Furthermore, HFCS adds longevity to the food's shelf life. However, over consumption of HFCS has been associated to poor health outcomes.[10, 11]

Disaccharides are carbohydrates composed of two monosaccharides units that are linked through a glycosidic bond. The most common disaccharides found in food include sucrose, lactose, and maltose. Sucrose is the most common disaccharide composed of a Glc ($\beta1\rightarrow2$) Fru primarily used as sweetener or table sugar. The non-reducing saccharide sucrose is naturally found in fruits, honey, maple sugar, and vegetables in lesser amounts. This disaccharide is typically extracted from sugar cane plants and sugar beets mass-produced for commercial use in the food industry. Lactose, present in many dairy products, is composed of Gal ($\beta1\rightarrow4$) Glc and makes up to 2-8% by weight in milk.[12] It is also commonly added to make milk and baked products.[13] Lastly, maltose is a disaccharide composed of Glc ($\alpha1\rightarrow4$) Glc and is less common than sucrose and lactose. Maltose is composed of two glucose monosaccharides linked at the $\alpha1\rightarrow4$ position and is a common byproduct of starch degradation using alpha-amylases.[14]

Plant-based foods also contain native oligosaccharides. Fructooligosaccharides (FOS), galactooligosaccharides (GOS), xylooligosaccharides (XOS) are examples of different oligosaccharide families found in plants. A common FOS found in legumes is a class of raffinose family oligosaccharides (RFO).[15] Raffinose is a hexose trisaccharide (Hex_3) composed of a Gal ($\alpha1\rightarrow6$) Glc ($\alpha1\rightarrow2$) Fru structure. Other RFO includes stachyose, a Hex_4 oligomer, and verbascose, a Hex_5 oligomer.[16] These oligosaccharides are produced in the vacuole of plant cells and are found in beans, cabbage, broccoli, and other vegetables.[17] While these compounds are

not digested in the stomach and small intestine, they can be readily fermented in the large intestine by gut bacteria producing beneficial organic acids and even flatulence.[18, 19]

Most food polysaccharides are found in plant and fungal cell walls and exhibit an enormous amount of structural diversity. Homopolysaccharides are those compounds that contain repeating units of the same monosaccharides and have the same repeating anomer content. For example, galactan, a side chain of the pectin backbone, consists of repeating (β1→4) Gal units. Heteropolysaccharides are those that contain multiple monosaccharides, diverse glycosidic linkages, and spatial features. For example, a heteropolysaccharide, galactomannan, comprises a repeating β1→4 mannan backbone capped with (α1→6) Gal units approximately every 4-6 mannan residues. Furthermore, galactomannan from various sources can have different galactose substitution patterns. For example, galactomannan from guar is 38 % substituted with galactose, and galactomannan from carob seed is only 22% substituted with galactose.[20] Often polysaccharides from different plant species and tissue have different structures.[21, 22] Common cell wall polysaccharides found in plant cells and thus food as part of a complex and heterogeneous mixture are displayed in **Figure 1.2**. In addition, plants and foods contain different amounts of these cell wall polysaccharides. For example, fruits contain higher amounts of pectin polysaccharides, whereas, in grains, arabinoxylans are the most abundant polymers aside from starches.[23, 24] **Table 1.2** illustrates some unique cell wall polysaccharides found in foods.[25]
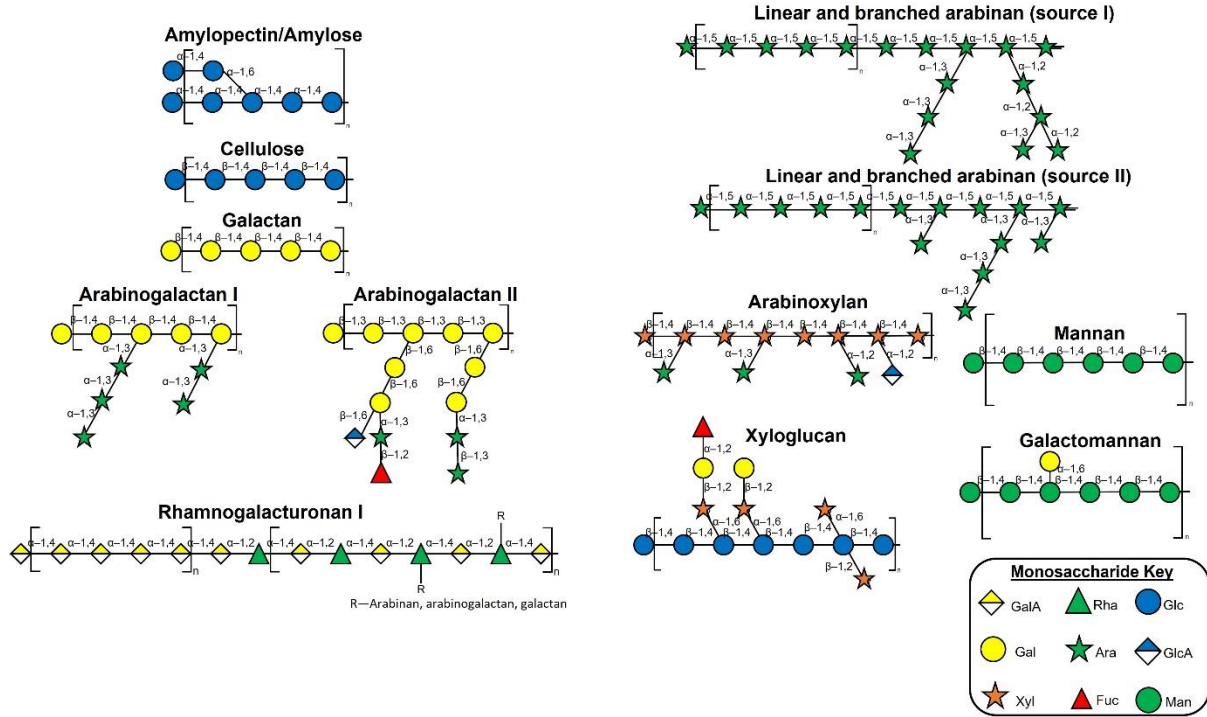
Figure 1.2. Polysaccharide structures commonly found in plant cells.

Table 1.2 List of polysaccharides and oligosaccharides found in plant cells and major food sources. Table reprinted with permission from *Critical Reviews in Food science and Nutrition.*[25]

Class	Polymer	Structure	Major sources
Starch	Amylose	$(1\rightarrow4)$-α-linked D-glucose	Cereals, tubers, legumes, pulses
	Amylopectin	$(1\rightarrow4)$-α-linked D-glucose	
		$(1\rightarrow6)$-α-linked branches	
Glucose	Cellulose	$(1\rightarrow4)$-β-linked D-glucose	Fruit, vegetables
	Callose	$(1\rightarrow3)$-β-linked D-glucose	Cereal grains
	Mixed linkage glucan	$(1\rightarrow3,1\rightarrow4)$-$\beta$-linked D-glucose	Seaweeds, yeast and other fungi
	Mixed linkage glucan	$(1\rightarrow3,1\rightarrow6)$-$\beta$-linked D-glucose	
Hemicellulose	Xyloglucan	$(1\rightarrow4)$-β-linked D-glucose	Fruit, vegetables, tamarind
	Glucomannan	$(1\rightarrow6)$-α-linked D-xylose substitutions	
		$(1\rightarrow4)$-β-linked D-mannose	
		$(1\rightarrow6)$-α-linked D-glucose substitutions	
	Galactomannan	$(1\rightarrow4)$-β-linked D-mannose	Guar, locust, and carob beans
	Glucuronomannans	$(1\rightarrow6)$-α-linked D-galactcose substitutions	Fungi, algae
	Galactans	$(1\rightarrow2)$-β-linked D-mannose and D-glucuronic acid	Seaweeds (carageenins, agar)
	Arabinoxylan	D-galactose and L-arabinose substitutions	Cereal grain
		$(1\rightarrow3)$-β-galactose	Cereal grain
		$(1\rightarrow4)$-3,6-anhydro-α-D or L-galactose	
		$(1\rightarrow4)$-β-xylose$(1\rightarrow2)$-α and	
		$(1\rightarrow3)$-α-L-arabinose and ferulyolated L-arabinose substitutions	
	Glucoronarabinoxylans	As arabinoxylans with D-glucuronic acid substitutions	
Pectins	Homogalacturan (HG)	Highly methyl esterified chains of	Fruit and vegetables
		$(1\rightarrow4)$-α-D-galacturonic acid	Fruit and vegetables
	(RGI) Rhamnogalacturan I (RGI)	Repeated $(1\rightarrow4)$-α-linked D-galacturonic acid	Fruit and vegetables
		$(1\rightarrow2)$-α-D-rhamnose disaccharides.	Fruit and vegetables
		Substitutions of rhamnose with	
		$(1\rightarrow4)$-β-galactan, arabinan, arabinogalactan chains	
	Rhamnogalacturan II (RG II)	HG backbone with side chains containing several types of sugar linkage	
Oligosaccharides	Fructans	$(1\rightarrow2)$-β-linked-D-fructose	Chicory, Jerusalem artichoke, cereals
	Raffinose	D-galactose $(1\rightarrow6)$-α-D-glucose $(1\rightarrow2)$-β-D-fructose	Legumes, vegetables
	Stachyose	Galactose $(1\rightarrow6)$-α-raffinose	Legumes, vegetables

Function

Carbohydrates in plant cells have many functions. One function of the cell wall polysaccharides is to provide structural rigidity to the cell and plant tissues. The plant cell wall comprises cellulose microfibrils intertwined with xyloglucans and rhamnogalacturonan I &II. Cellulose, the primary wall component, provides cellular structural support.[21] In contrast, secondary wall components such as xyloglucan and rhamnogalacturonan provide elasticity and modularity to the cell wall.[26] Furthermore, the secondary cell wall components further provides water impermeability capabilities and prevents cellulose aggregation to facilitate cell wall

expansion.[27] A model of the cell wall in a *Arabidopsis* leaf tissue with three major classes of

polysaccharides: cellulose, hemicelluloses and pectins is shown in **Figure 1.3**.
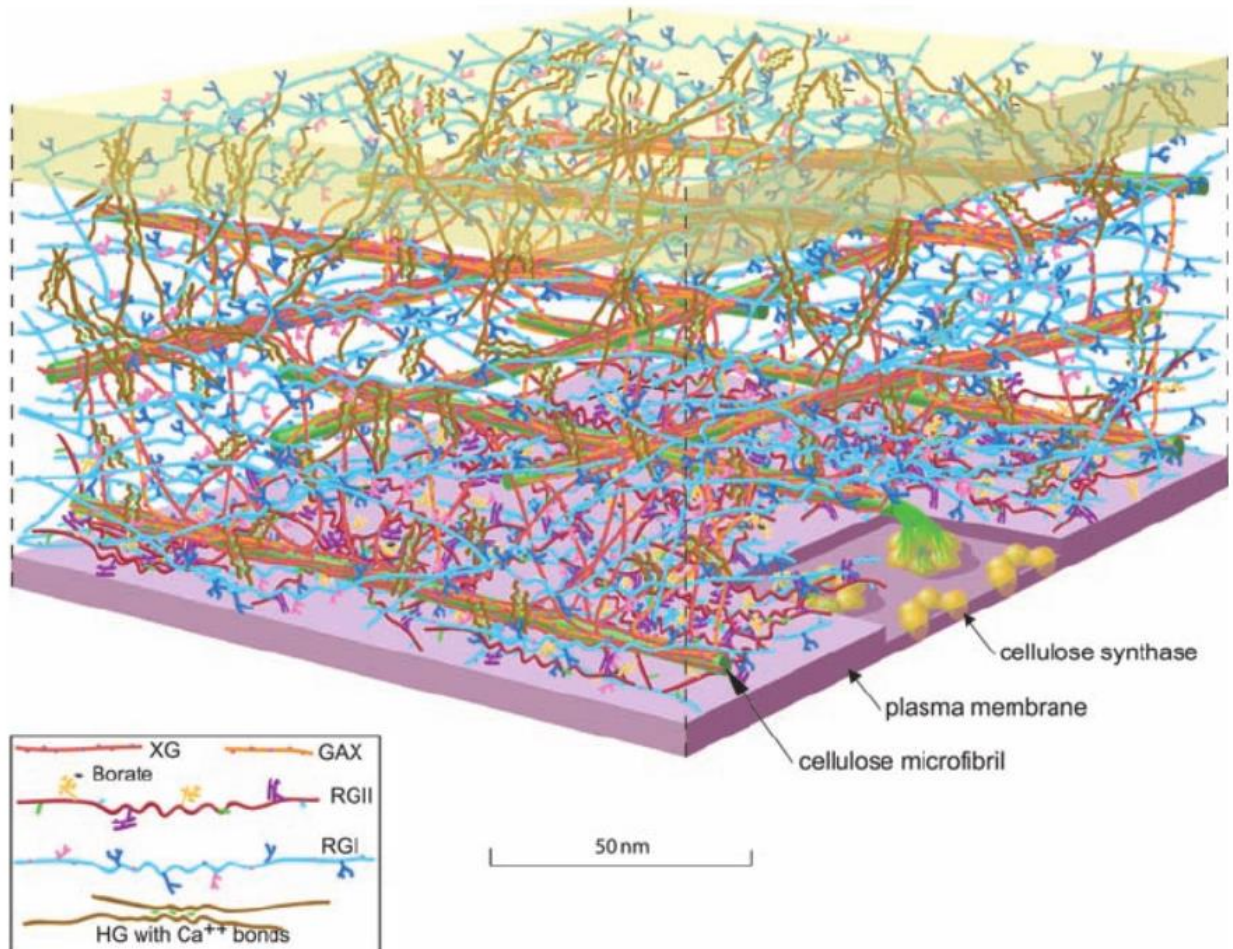
Figure 1.3. Plant cell wall illustrates polysaccharide diversity, composition, and complexity.

Secondary cell wall components such as xyloglucan (XG), glucoronoaraninoxylan (GAX),

homogalacturonon (HG), and rhamnogalacturonoan I &II (RGI and RGII) are all intertwined

around cellulose microfibrils. Figure reprinted with permission from *Science*.[21]

Another essential function of polysaccharides is the energy source they provide. [28] Starch

polysaccharides store energy in different regions of the plant (tissues, seed, and roots) in the

form of starch granules.[29] Starch is composed of two types of large molecular weight polysaccharides: amylose and amylopectin. Amylose is a long linear homopolysaccharide composed of a glucose backbone with ($\alpha1\rightarrow4$) glycosidic linkages. Amylopectin also has a glucose ($\alpha1\rightarrow4$) backbone, but has occasional branching points at ($\alpha1\rightarrow6$) linkage positions every 24-30 glucose residues.[30] The branched points in amylopectin increase accessibility for α amylases and amyloglucosidases to depolymerize the polysaccharides. In contrast, amylose has no branch points and can hydrogen bond on itself, making it more difficult for chemicals and enzymes to depolymerize the structure.

Biosynthesis

Carbohydrates are produced in plant cells by the process of photosynthesis. The biosynthesis of cell wall polysaccharides is a highly sophisticated process that uses multiple enzymes, and many metabolic intermediates.[31] Glycosyltransferases (GTs) are responsible for forming glycosidic linkages from activated saccharides or nucleotide sugars. Uracil-diphosphate (UDP)-D-glucose is the initial precursor to synthesize the cellulose and micro-fibril polymers. These polymers are then extruded to the outer plasma membrane for structural support of the cell. Furthermore, other nucleotide sugar precursors like UDP-D-galactose, UDP-L-rhamnose, UDP-D-glucoronic acid, UDP-D-xylose, UDP-L-arbinose, guanosine diphosphate (GDP)-D-glucose, and adenosine diphosphate (ADP)-glucose are further converted in the Golgi into different oligosaccharides and polysaccharides.[32-34] The biosynthetic pathway for polysaccharides in *Arabidopsis* tissues is shown in **Figure 1.3**. Secondary cell wall polysaccharides such as xylan and glucomannans synthesis require more enzymes and are substantially more complicated.[35]
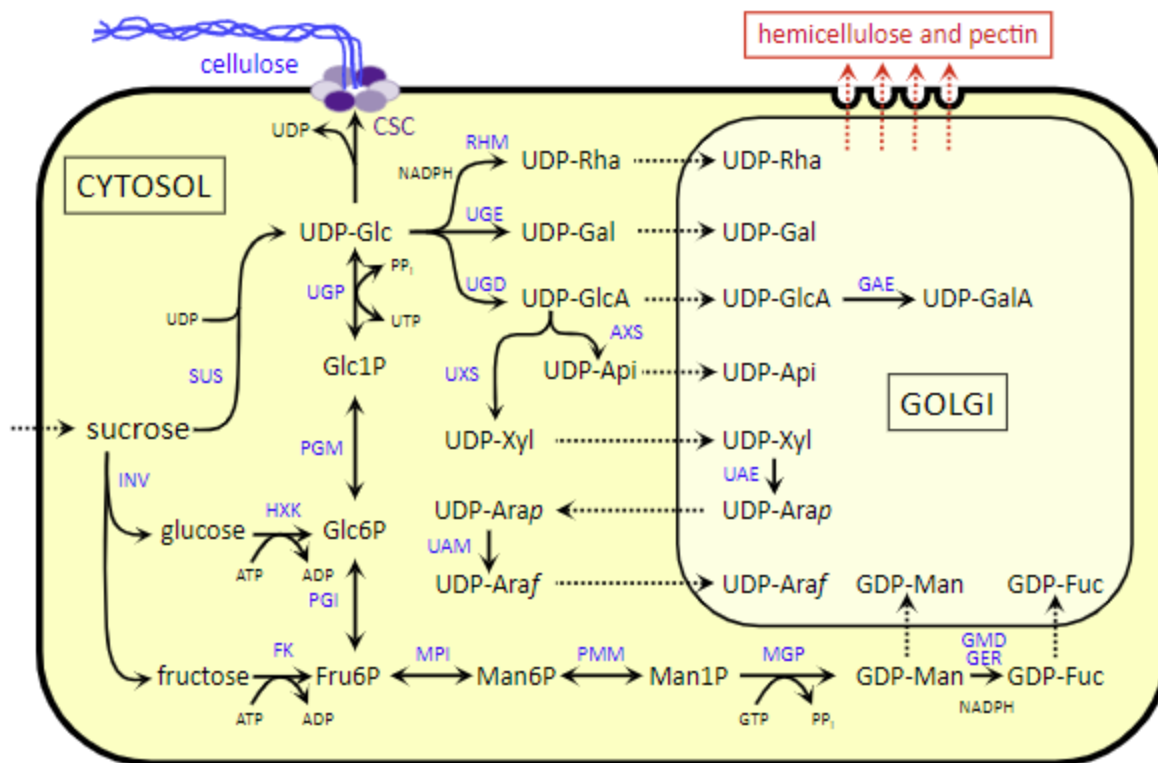
Figure 1.3. Biosynthesis process to produce cell wall polysaccharides Figure reprinted with permission from *Molecular Plant*.[31]

Carbohydrates as Prebiotics

Plant cell wall polysaccharides also serve an important nutritional role for humans. Nearly every cell wall polysaccharide is indigestible by human digestive enzymes and can thus serve as food for microbial communities found in the colon.[36] These carbohydrates are often termed "dietary fibers" and can reach the large intestine where they are available for degradation and utilization by host gut microbes. Dietary fiber can be termed "prebiotic," in which it is defined as "a substrate that is selectively utilized by host microorganisms conferring a health benefit."[37] In recent years, there has been increasing interest in utilizing the gut microbiome to modulate our health. The detailed knowledge of carbohydrate-microbe interactions remains

11

limited. In general, however, the fermentation of dietary fiber results in the production of

metabolites as short chain-fatty acids (SCFA). These metabolites provide the host with benefits

such as immune support and are associated with reducing the risk of and ameliorating various

metabolic disorders.[38-40]

The human gut is home to trillions of bacteria primarily found in the large intestine.[41]

These bacterial species collectively possess thousands of genes encoding for the production of

carbohydrate-active enzymes (CAZymes) that used to breakdown and utilize the diverse

structures within dietary fiber.[42] **Table 1.3** displays list of dietary fibers and some corresponding

glycosyl hydrolyses (GH) and polysaccharide lyases (PL) to degrade respective substrate.

Table 1.3. Illustrates 18 dietary fibers with their respective chemical structure. Each

carbohydrate substrate has unique degradation enzymes and CAZymes families GHs. Table

reprinted with permission from *Journal of Molecular Biology* .[43]

Dietary fibers	Chemical structures	Degradation enzymes[a]	CAZymes families[b]
RS	Linear α-(1,4) Glu units, branched with α-(1,6) Glu	α-Amylase, glucoamylase, isoamylase, pullulanase,	GH4, GH13, GH14, GH15, GH31, GH57, GH63, GH97
Pectins[c]			
Homogalacturonan	Linear α-(1,4) GalA units partially methylated and acetylated	Polygalacturonase, pectin methyl esterase, pectin lyase, pectin acetyl esterase	GH28, CE8, CE12, CE13, PL1, PL9, PL10
RG I	Linear α-(1,4) GalA and α-(1,2) Rha units	RG lyase, RG hydrolase, RG acetyl esterase	GH2, GH28, GH42, GH105, CE4, CE6, CE12, PL4, PL9, PL11
Pectic galactan	Composed of β-(1,4) Gal, β-(1,6) Gal, α-(1,4) Ara units	β-(1,4) Galactanase, β-(1,6) galactanase, α-arabinosidase	GH2, GH3, GH5, GH10, GH30, GH43, GH53
Arabinogalactan	Composed of β-(1,3) Gal, β-(1,6) Gal, α-(1,5) Ara, and (1,3) Ara units	β-(1,3) Galactanase, β-(1,6) galactanase, α-(1,5) arabinanase, α-L-arabinofuranosidase	GH2, GH3, GH10, GH30, GH35, GH43, GH51
Arabinan	Composed of α-(1,5), (1,3) and α-(1,2) Ara units	α-(1,5) Arabinanase, α-L-arabinofuranosidase	GH2, GH3, GH10, GH43, GH51, GH54, GH62
RG II[d]	Linear α-(1,4) GalA units branched with Api, Ara, AceA, Dha, Fuc, Gal, GluA, Kdo, Rha, Xyl	Not known[d]	
Cellulose	Linear β-(1,4) Glu units	Endoglucanase, cellulobiohydrolases, β-glucosidases	GH1, GH3, GH5, GH7, GH8, GH9, GH44, GH48, GH51, GH74
Hemicelluloses			
Arabinoxylan[c]	β-(1,4) Xyl units as the backbone with side chains of Ara units linked via α-(1,2), α-(1,3), and α-(1,5). Gal, GluA, and FerA units may also be present in the branched points	Xylanase, α-glucuronidase, α-L-arabinofuranosidase, β-xylosidase, feruloyl esterase, acetyl xylan esterase	GH3, GH5, GH7, GH8, GH10, GH11, GH39, GH43, GH51, GH52, GH54, GH62, GH67, GH115, CE1, CE2, CE4, CE6, CE7
Xyloglucan	β-(1,4) Glu units as the backbone with single unit side chains of α-(1,6) Xyl. Gal, and Fuc may also present in the branched point	Xylogluconase, xylosidase, β-D-gluconase, β-D-galactosidase	GH1, GH2, GH5, GH7, GH12, GH16, GH42, GH44, GH45, GH48, GH51, GH74
Glucomannan	Linear or slightly branched backbone chain of β-(1,4) Man and β-(1,4) Glu	β-Mannanase, β-mannosidase, β-glucosidase, α-galactosidase	GH1, GH2, GH4, GH5, GH26, GH27, GH36, GH57, GH97, GH110, GH113
Galactomannan	Linear or slightly branched backbone chain of β-(1,4) Man and β-(1,4) Glu with side chains of α-(1,6) Gal	β-Mannanase, β-mannosidase, β-glucosidase, α-galactosidase	GH1, GH2, GH4, GH5, GH26, GH27, GH36, GH57, GH97, GH110, GH113.
β- Glucan	Repeating linear polymer of two β-(1,4) Glu alternated with β-(1,3) Glu units	Licheninase, β-glucan endohydrolase, endo (1,4) β-glucanase	GH5, GH6, GH8, GH9, GH10, GH12, GH16, GH26, GH44, GH45, GH48, GH51
Polyfructans			
Inulin	β-(2,1) Fru	Inulinase, β-(2,1) fructanlyase	GH32, GH91
Levan	β-(2,6) Fru	Levanase, β-fructofuronosidase	GH32, GH68
Gums			
Carrageenan	Sulfated galactans, units of β-(1,3) Gal and α-(1,4) linked 3,6 anhydro Gal	κ-Carrageenase	GH16
Alginate	Linear β-(1,4) ManA and α-(1,4) GluA	Alginate lyase	PL5, PL6, PL7, PL15, PL17

Abbreviations used: AceA, aceric acid; Ara, arabinose; Api, apiose; Dha, 3-deoxy-D-lyxo-2-heptulosaric acid; FerA, ferulic acid; Fru, fructose; Fuc, fucose; Gal, galactose; GalA, galacturonic acid; Glu, glucose; GluA, glucuronic acid; Kdo, 3-deoxy-D-manno-2-octulosonic acid; Man, mannose; Man, mannuronic acid; Rha, rhamnose; Xyl, xylose.

[a] Enzymes required for complete degradation of the corresponding dietary fiber structure.

[b] Possible CAZymes families that harbor the corresponding degradation enzymes for the given dietary fiber structure. For more detail, refer to CAZymes database (www.cazy.org) [24].

[c] Chemical structures of pectin and arabinoxylans are illustrated below.

[d] Enzymes degrading RG II are not known [88]; however, Martens *et al.* reported that enzymes secreted by *B. thetaiotaomicron* to utilize RG II belong to enzymes families of GH2, GH28, GH33, GH43, GH78, GH95, GH105, GH106, and PL1 [75].

Carbohydrate Analysis

Dietary carbohydrates derived from plants are highly challenging to characterize structurally. Their structural complexity arises from their diverse monosaccharide and glycosidic linkage compositions, disordered sequences, and anomer content. Furthermore, tertiary and quaternary structural dynamics add another layer of difficulty to the analysis and make chemical modifications challenging to employ. As a result, plant carbohydrates are far more complicated, diverse, and larger than those found in animals. In addition, polysaccharides can have masses up to a million Daltons in size, which provides the need for modification for their analysis. Furthermore, their structures contain a variety of isomers, each with distinct anomeric content, glycosidic linkage, and spatial structure. Often, cell wall polysaccharides contain modifications such as methylated species or acetyl groups attached to specific monosaccharides, which are difficult to capture.

For years the analysis of carbohydrates entailed using traditional analytical methods for characterization. The traditional carbohydrate analysis utilized a gas chromatography-mass spectrometry (GC-MS) approach.[44] The sample preparation for these analyses involves an extensive sample preparation that takes up to three days and requires many derivatization steps.[45, 46] Furthermore, the chromatography separation on GC has long run times and can very well reach over an hour. In addition, the samples are prepared and analyzed in a single sample manner, thereby providing very low throughput. Lastly, these analyses provide minimal structural specificity, limited carbohydrate monitoring, and do not provide absolute quantitation. Hence, there is a need for modern analytical methods to characterize carbohydrates.

15

Liquid chromatography coupled to mass spectrometry (LC-MS) is another alternative for carbohydrates analysis. The LC-MS analysis provides excellent figures of merit, including high sensitivity, high selectivity, high reproducibility, high mass accuracy (often linked to a high mass resolution instrument), and capable of high throughput analysis. However, the analysis of carbohydrates has not advanced as other biomolecules, such as proteomics and genomics. On the other hand, although amendable to LC-MS, carbohydrates exhibit many unique challenges and usually require multiple analytical approaches for their characterization. The slow progression in the field has severely limited the studies of these important molecules. Due to recent interest in the gut microbiome and its role in modulating our health, there is a need to characterize carbohydrates in food as they are the main driving force for shaping gut microbial communities.

Monosaccharide analysis

One approach to characterize food carbohydrates is monosaccharide compositional analysis using an LC-MS platform. This strategy involves the hydrolysis of disaccharides, oligosaccharides, and polysaccharides to their respective monosaccharide units.[47] The liberated monosaccharides are then derivatized with 1-phenyl-3-methyl-5-pyrazolone (PMP) followed by a C18 chromatographic separation (**Figure 1.4b**) and subjected to a targeted dynamic multiple reaction monitoring (dMRM) MS analysis.[48] Two product ions from the PMP labeled monosaccharide are monitored, enabling further confirmation of the monosaccharide structure. The most abundant product ion *m/z* 175.0 is used as a quantifying ion. The qualifier ions monitored are *m/z* 217.2 and *m/z* 216.1 which can distinguish between hexoses and N-acetylglucosamine, respectively (**Figure 1.5**). This analysis provided simultaneous monitoring of 14 monosaccharides, and the results yielded absolute quantitation using an external calibration

curve. The limit of detection approached the attomole level while providing an excellent linear

dynamic range of up to 6 orders of magnitude.
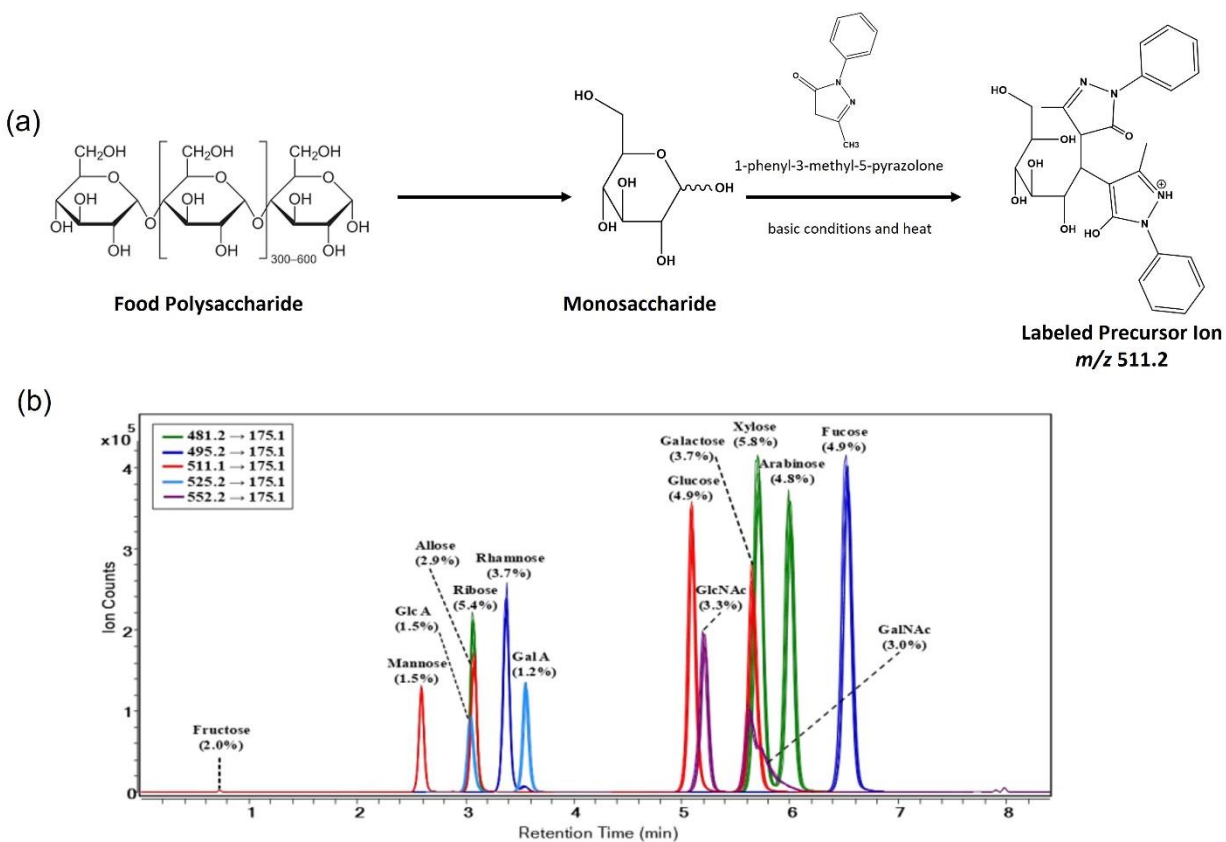
Figure 1.4. Monosaccharide compositional analysis sample workflow (a) and dMRM

chromatogram of 14 plant monosaccharides (b). Figure reprinted with permission from

International Journal of Mass Spectrometry.[47]
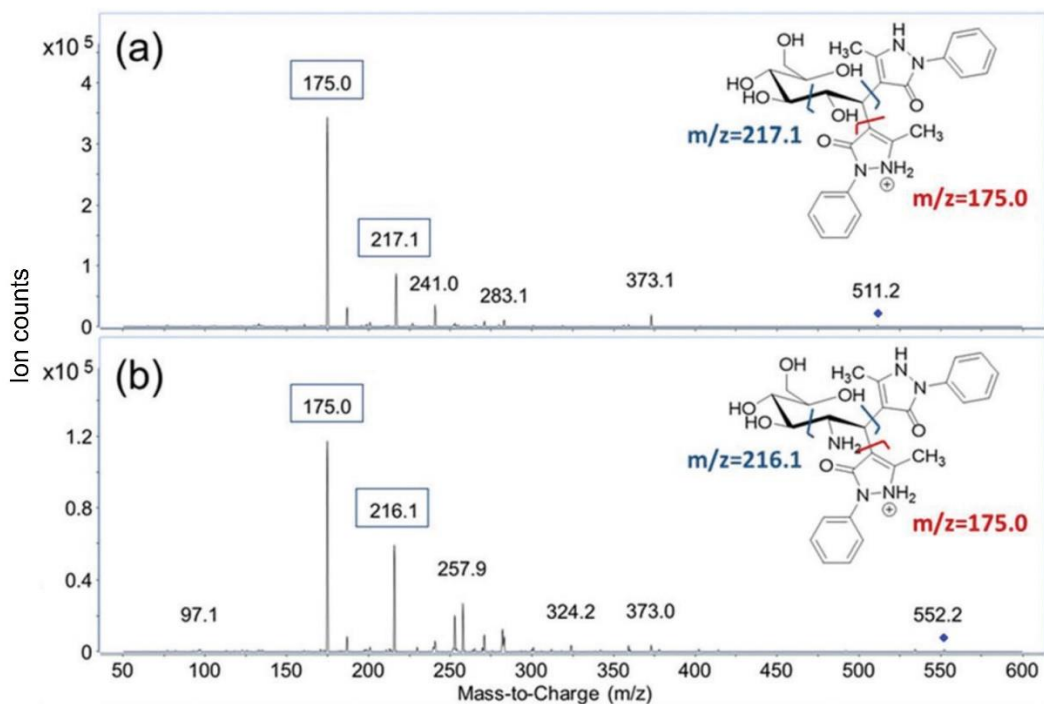
Figure 1.5. MS/MS spectra of PMP labeled glucose (a), and PMP labeled GlcNAc (b). Figure reprinted with permission from *Analyst*.[48]

Linkage analysis

Another approach to characterize plant carbohydrates is glycosidic linkage analysis. The linkage analysis provides unique structural information of the parent carbohydrate. A linkage can be determined from a compound containing two or more monosaccharides linked to the glycosidic bond. The method entails complete methylation of all free hydroxyl groups using saturated NaOH in DMSO and iodomethane. After complete methylation, a hydrolysis step is employed to release saccharides at the glycosidic bond, followed by PMP labeling using the protocol as in the monosaccharide analysis. The derivatized glycosides are subjected to UHPLC separation and detection using an MRM MS method. The recently reported linkage analysis method is the first analysis performed on an LC-MS platform.[49] The linkage profile of standard

oligosaccharides is shown in **Figure 1.6**. There are serval advantages to using the LC-MS

approach over the traditional GC-MS approach; namely, this approach provides more detected

linkages, greater sensitivity, higher throughput and only requires 50 µg of starting material for

analysis.
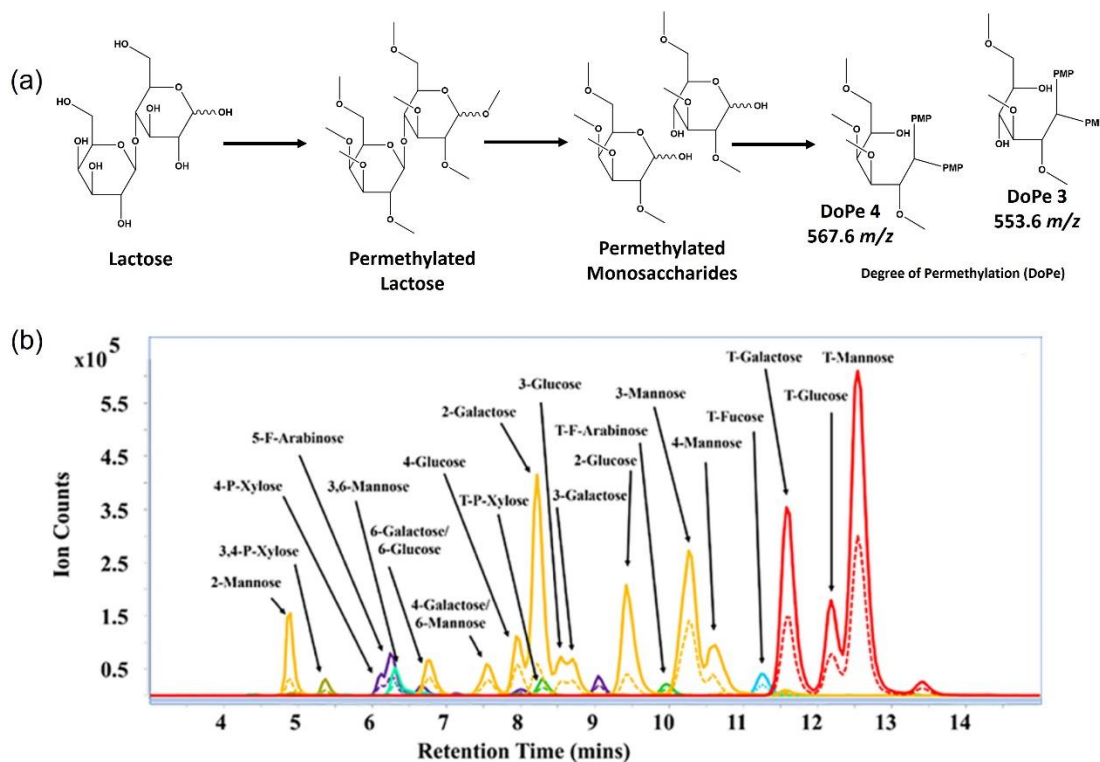
Figure 1.6. Complete glycosidic linkage analysis workflow on an LC-MS platform (a) and the

MRM chromatogram of the linkage present in a pool of known oligosaccharide structures (b).

Figure reprinted with permission from *Analytical Chemistry*.[49]

Sequence analysis

Lastly, methods for oligosaccharide analysis involving liquid chromatography-tandem

mass spectrometry (LC-MS/MS) have been developed for the structural elucidation of

carbohydrates.[50-53] This method involves fragmenting the precursor ion by collision induced

dissociation (CID) and monitoring generated product ions. The LC-MS/MS platform yields

structural information such as degree of polymerization (DP), monosaccharides classes, and

sequence, revealing the connectivity of the oligosaccharide structure.[54] Furthermore, a novel

middle-up approach can be employed to characterize polysaccharides. However, due to the large

size of polysaccharides, their direct analysis by MS is not feasible and must be depolymerized

into respective oligosaccharides. Typically, mild acid hydrolysis methods and enzymatic

approaches are used for their depolymerization. However, they have many disadvantages such as

producing monosaccharides and requiring multiple enzymes to depolymerize all food

polysaccharides. Therefore, a new method to depolymerize polysaccharides was recently

developed.[55] The method involves a Fenton's chemistry where a Fe^{3+} catalyst reacts with H_2O_2

to generate hydroxy radials and induce oxidative cleavage on the backbone of polysaccharides to

create oligosaccharides. The oligosaccharides generated can then be probed for LC-MS analysis.
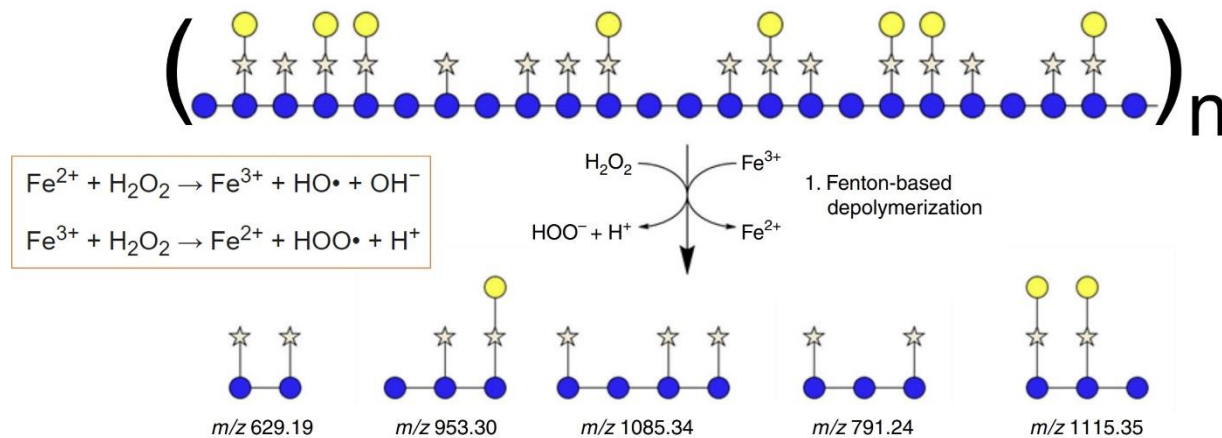
Figure 1.7. Fenton's reaction to oxidatively depolymerize xyloglucan into respective

oligosaccharides. The released oligosaccharides can further be probed on an LC-MS platform for

profiling.

CONCLUSION

Dietary carbohydrates are the most abundant components in foods. These have many biological functions, such as providing energy, structural support for cells and plant tissues, and a major food source for gut bacteria. These biomolecules have essential biological functions, and are the main driving force for modulating the gut microbiome. However, carbohydrates are difficult to structurally characterize. Traditional carbohydrate analysis are performed on a GC-MS platform; however, the modern LC-MS approach has revitalized carbohydrate analysis with better figures of merit (throughput, sensitivity, and specificity). The chapters herein encompass a series of LC-MS-based methods for the structural characterization of dietary carbohydrates. The results from the work presented will enable a more profound understanding of the interaction between carbohydrate structures revealing the unknown mystery of carbohydrates compositions in foods.

REFERENCES

1.	Harvey, D. J.; Merry, A. H.; Royle, L.; Campbell, M. P.; Rudd, P. M., Symbol nomenclature for representing glycan structures: Extension to cover different carbohydrate types. *Proteomics* **2011,** *11* (22), 4291-4295.

2.	Stanley, P.; Schachter, H.; Taniguchi, N.; Varki, A.; Cummings, R.; Esko, J.; Freeze, H.; Bertozzi, C.; Hart, G.; Etzler, M., Essentials of glycobiology. 2009.

3.	Cummings, J.; Stephen, A., Carbohydrate terminology and classification. *European Journal of Clinical Nutrition* **2007,** *61* (1), S5-S18.

4.	Navarro, D. M.; Abelilla, J. J.; Stein, H. H., Structures and characteristics of carbohydrates in diets fed to pigs: a review. *Journal of Animal Science and Biotechnology* **2019,** *10* (1), 1-17.

5.	Lunn, J.; Buttriss, J., Carbohydrates and dietary fibre. *Nutrition Bulletin* **2007,** *32* (1), 21-64.

6.	Davidson, R. L., *Handbook of water-soluble gums and resins*. 1980.

7.	Christianson, D.; Hodge, J.; Osborne, D.; Detroy, R. W., Gelatinization of wheat starch as modified by xanthan gum, guar gum, and cellulose gum. **1981**, *58* (6), 513-517.

8.	Sanderson, G. R., Applications of xanthan gum. *British Polymer Journal* **1981,** *13* (2), 71-75.

9.	Marx, J. T.; Marx, B. D.; Johnson, J. M., High fructose corn syrup cakes made with all purpose flour or cake flour. *Cereal Chemistry* **1990,** *67* (5), 502-4.

10.	Bray, G. A.; Nielsen, S. J.; Popkin, B. M., Consumption of high-fructose corn syrup in beverages may play a role in the epidemic of obesity. *The American Journal of Clinical Nutrition* **2004,** *79* (4), 537-543.

11.	Ferder, L.; Ferder, M. D.; Inserra, F., The role of high-fructose corn syrup in metabolic syndrome and hypertension. *Current Hypertension Reports* **2010,** *12* (2), 105-112.

12.	Cerbulis, J.; Farrell Jr, H., Composition of milks of dairy cattle. I. Protein, lactose, and fat contents and distribution of protein fraction. *Journal of Dairy Science* **1975,** *58* (6), 817-827.

13.	Langemeier, J.; Rogers, D., Rapid method for sugar analysis of doughs and baked products. *Cereal Chemistry (USA)* **1995**, *72* (4), 349-351.

14.	Lentze, M. J., The history of maltose-active disaccharidases. *Journal of Pediatric Gastroenterology and Nutrition* **2018,** *66*, S4-S6.

15.	Sosulski, F.; Elkowicz, L.; Reichert, R., Oligosaccharides in eleven legumes and their air-classified protein and starch fractions. *Journal of Food Science* **1982,** *47* (2), 498-502.

16.	ElSayed, A. I.; Rafudeen, M. S.; Golldack, D., Physiological aspects of raffinose family oligosaccharides in plants: protection against abiotic stress. *Plant Biology* **2014,** *16* (1), 1-8.

17.	Jovanovic-Malinovska, R.; Kuzmanova, S.; Winkelhausen, E., Oligosaccharide profile in fruits and vegetables as sources of prebiotics and functional foods. *International Journal of Food Properties* **2014,** *17* (5), 949-965.

18.	Mano, M. C. R.; Neri-Numa, I. A.; da Silva, J. B.; Paulino, B. N.; Pessoa, M. G.; Pastore, G. M., Oligosaccharide biotechnology: an approach of prebiotic revolution on the industry. *Applied Microbiology and Biotechnology* **2018,** *102* (1), 17-37.

19.   Kovács, Z.; Benjamins, E.; Grau, K.; Rehman, A. U.; Ebrahimi, M.; Czermak, P., Recent developments in manufacturing oligosaccharides with prebiotic functions. *Biotechnology of Food and Feed Additives* **2013**, 257-295.

20.   Srivastava, M.; Kapoor, V., Seed galactomannans: an overview. *Chemistry & Biodiversity* **2005,** *2* (3), 295-317.

21.   Somerville, C.; Bauer, S.; Brininstool, G.; Facette, M.; Hamann, T.; Milne, J.; Osborne, E.; Paredez, A.; Persson, S.; Raab, T., Toward a systems approach to understanding plant cell walls. *Science* **2004,** *306* (5705), 2206-2211.

22.   Zhang, Q.; Cheetamun, R.; Dhugga, K. S.; Rafalski, J. A.; Tingey, S. V.; Shirley, N. J.; Taylor, J.; Hayes, K.; Beatty, M.; Bacic, A., Spatial gradients in cell wall composition and transcriptional profiles along elongating maize internodes. *BMC Plant Biology* **2014,** *14* (1), 1-19.

23.   Baker, R. A., Reassessment of some fruit and vegetable pectin levels. *Journal of Food Science* **1997,** *62* (2), 225-229.

24.   Bengtsson, S.; Andersson, R.; Westerlund, E.; Åman, P., Content, structure and viscosity of soluble arabinoxylans in rye grain from several countries. *Journal of the Science of Food and Agriculture* **1992,** *58* (3), 331-337.

25.   Lovegrove, A.; Edwards, C.; De Noni, I.; Patel, H.; El, S.; Grassby, T.; Zielke, C.; Ulmius, M.; Nilsson, L.; Butterworth, P., Role of polysaccharides in food, digestion, and health. *Critical Reviews in Food science and Nutrition* **2017,** *57* (2), 237-253.

26.   Somerville, C., Cellulose synthesis in higher plants. *Annu. Rev. Cell Dev. Biol.* **2006,** *22*, 53-78.

27.   Chanliaud, E.; De Silva, J.; Strongitharm, B.; Jeronimidis, G.; Gidley, M. J., Mechanical effects of plant cell wall enzymes on cellulose/xyloglucan composites. *The Plant Journal* **2004,** *38* (1), 27-37.

28.   Pérez, S.; Baldwin, P. M.; Gallant, D. J., Structural features of starch granules I. In *Starch*, Elsevier: 2009; pp 149-192.

29.   French, D., Organization of starch granules. In *Starch: Chemistry and Technology*, Elsevier: 1984; pp 183-247.

30.   Hirst, E.; Plant, M.; Wilkinson, M., 344. Polysaccharides. Part XIV. The molecular structure of amylose and of amylopectin. *Journal of the Chemical Society (Resumed)* **1932**, 2375-2383.

31.   Verbančič, J.; Lunn, J. E.; Stitt, M.; Persson, S., Carbon supply and the regulation of cell wall synthesis. *Molecular Plant* **2018,** *11* (1), 75-94.

32.   Leloir, L. F., Nucleoside diphosphate sugars and saccharide synthesis. The fourth Hopkins Memorial Lecture. *Biochemical Journal* **1964,** *91* (1), 1. b2.

33.   Hassid, W., Biosynthesis of Oligosaccharides and Polysaccharides in Plants: Mechanisms of enzymic synthesis of complex plant carbohydrates are reviewed. *Science* **1969,** *165* (3889), 137-144.

34.   Seifert, G. J., Nucleotide sugar interconversions and cell wall biosynthesis: how to bring the inside to the outside. *Current Opinion in Plant Biology* **2004,** *7* (3), 277-284.

35.   Zhong, R.; Cui, D.; Ye, Z. H., Secondary cell wall biosynthesis. *New Phytologist* **2019,** *221* (4), 1703-1723.

36.   Flint, H. J.; Bayer, E. A.; Rincon, M. T.; Lamed, R.; White, B. A., Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. *Nature Reviews Microbiology* **2008,** *6* (2), 121-131.

37.   Gibson, G. R.; Hutkins, R.; Sanders, M. E.; Prescott, S. L.; Reimer, R. A.; Salminen, S. J.; Scott, K.; Stanton, C.; Swanson, K. S.; Cani, P. D., Expert consensus document: The International Scientific Association for Probiotics and Prebiotics (ISAPP) consensus statement on the definition and scope of prebiotics. *Nature Reviews Gastroenterology & Hepatology* **2017,** *14* (8), 491-502.

38.   Topping, D. L.; Clifton, P. M., Short-chain fatty acids and human colonic function: roles of resistant starch and nonstarch polysaccharides. *Physiological Reviews* **2001**.

39.   De Goffau, M. C.; Luopajärvi, K.; Knip, M.; Ilonen, J.; Ruohtula, T.; Härkönen, T.; Orivuori, L.; Hakala, S.; Welling, G. W.; Harmsen, H. J., Fecal microbiota composition differs between children with β-cell autoimmunity and those without. *Diabetes* **2013,** *62* (4), 1238-1244.

40.   Ley, R. E.; Turnbaugh, P. J.; Klein, S.; Gordon, J. I., Human gut microbes associated with obesity. *Nature* **2006,** *444* (7122), 1022-1023.

41.   Gill, S. R.; Pop, M.; DeBoy, R. T.; Eckburg, P. B.; Turnbaugh, P. J.; Samuel, B. S.; Gordon, J. I.; Relman, D. A.; Fraser-Liggett, C. M.; Nelson, K. E., Metagenomic analysis of the human distal gut microbiome. *Science* **2006,** *312* (5778), 1355-1359.

42.   Garron, M.-L.; Henrissat, B., The continuing expansion of CAZymes and their families. *Current opinion in chemical biology* **2019,** *53*, 82-87.

43.   Hamaker, B. R.; Tuncil, Y. E., A perspective on the complexity of dietary fiber structures and their potential effect on the gut microbiota. *Journal of Molecular Biology* **2014,** *426* (23), 3838-3850.

44.   Frank, H.; Neves, H. J. C. D.; Bayer, E., Gas chromatography of monosaccharides: formation of a single derivative for each aldose. *Journal of Chromatography A* **1981,** *207* (2), 213-220.

45.   Ciucanu, I.; Kerek, F., A simple and rapid method for the permethylation of carbohydrates. *Carbohydrate Research* **1984,** *131* (2), 209-217.

46.   Kamerling, J. P.; Gerwig, G. J.; Vliegenthart, J.; Clamp, J., Characterization by gas-liquid chromatography-mass spectrometry and proton-magnetic-resonance spectroscopy of pertrimethylsilyl methyl glycosides obtained in the methanolysis of glycoproteins and glycopeptides. *Biochemical Journal* **1975,** *151* (3), 491-495.

47.   Amicucci, M. J.; Galermo, A. G.; Nandita, E.; Vo, T.-T. T.; Liu, Y.; Lee, M.; Xu, G.; Lebrilla, C. B., A rapid-throughput adaptable method for determining the monosaccharide composition of polysaccharides. *International Journal of Mass Spectrometry* **2019,** *438*, 22-28.

48.   Xu, G.; Amicucci, M. J.; Cheng, Z.; Galermo, A. G.; Lebrilla, C. B., Revisiting monosaccharide analysis–quantitation of a comprehensive set of monosaccharides using dynamic multiple reaction monitoring. *Analyst* **2018,** *143* (1), 200-207.

49.   Galermo, A. G.; Nandita, E.; Barboza, M.; Amicucci, M. J.; Vo, T.-T. T.; Lebrilla, C. B., Liquid chromatography–tandem mass spectrometry approach for determining glycosidic linkages. *Analytical Chemistry* **2018,** *90* (21), 13073-13080.

50.   Li, H.; Bendiak, B.; Siems, W. F.; Gang, D. R.; Hill Jr, H. H., Carbohydrate structure characterization by tandem ion mobility mass spectrometry (IMMS) 2. *Analytical Chemistry* **2013,** *85* (5), 2760-2769.

51.   Kawasaki, N.; Ohta, M.; Hyuga, S.; Hyuga, M.; Hayakawa, T., Application of liquid chromatography/mass spectrometry and liquid chromatography with tandem mass spectrometry to the analysis of the site-specific carbohydrate heterogeneity in erythropoietin. *Analytical Biochemistry* **2000,** *285* (1), 82-91.

52.   Zaia, J.; Costello, C. E., Tandem mass spectrometry of sulfated heparin-like glycosaminoglycan oligosaccharides. *Analytical Chemistry* **2003,** *75* (10), 2445-2455.

53.   An, H. J.; Lebrilla, C. B., Structure elucidation of native N-and O-linked glycans by tandem mass spectrometry (tutorial). *Mass Spectrometry Reviews* **2011,** *30* (4), 560-578.

54.   Castillo, J. J.; Galermo, A. G.; Amicucci, M. J.; Nandita, E.; Couture, G.; Bacalzo, N.; Chen, Y.; Lebrilla, C. B., A Multidimensional Mass Spectrometry-Based Workflow for De Novo Structural Elucidation of Oligosaccharides from Polysaccharides. *Journal of the American Society for Mass Spectrometry* **2021,** *32* (8), 2175-2185.

55.   Amicucci, M. J.; Nandita, E.; Galermo, A. G.; Castillo, J. J.; Chen, S.; Park, D.; Smilowitz, J. T.; German, J. B.; Mills, D. A.; Lebrilla, C. B., A nonenzymatic method for cleaving polysaccharides to yield oligosaccharides for structural analysis. *Nature Communications* **2020,** *11* (1), 1-12.

Chapter 2

Glycopedia: A Glycan Encyclopedia for Food

ABSTRACT

The molecular complexity of the carbohydrates we consume has been deceptively oversimplified due to a lack of analytical methods that possess the throughput, sensitivity, and resolution required to provide quantitative structural information. Such information is becoming an integral part of understanding how specific glycan structures impact health through their interaction with the gut microbiome and host physiology. In this work, the primary goal was to catalogue the glycans present in complementary foods commonly consumed as the first foods after breast milk by toddler. The monosaccharide compositions of over 800 foods from diverse food groups including fruits, vegetables, grain products, beans/peas/legumes/nuts/seeds, sweets and beverages, animal products, and more were obtained and used to construct the Glycopedia, an open-access database that provides quantitative structural information on the carbohydrates in food. While many foods within the same group possessed similar compositions, hierarchical clustering analysis revealed similarities between different groups as well. The Glycopedia can be used to formulate diets rich in specific monosaccharide residues to provide a more targeted modulation of the gut microbiome, thereby opening the door for a new class of therapeutic diets.

27

INTRODUCTION

Carbohydrates make up the largest component of human diets, comprising up to 85% depending on geographic location and socioeconomic status.[1] These biomolecules play a profound role in shaping our gut microbial communities, the spectrum of microbial metabolites produced and the resulting impacts on our health. For example, in early life, human milk oligosaccharides (HMOs) play a large role in feeding *Bifidobacteria,* thereby shaping the infant's microbial communities and providing health benefits such as priming the immune system, strengthening the gut barrier, and blocking pathogens.[2] In adults, a high fat/high carbohydrate or "Western" diet has long been implicated in a variety of metabolic diseases such as cardiovascular diseases, type 2 diabetes, obesity, and gastrointestinal disorders.[3, 4] On the other hand, consumption of plant-based foods are associated with reducing the risks of those metabolic diseases.[5]

Recent research has emphasized the important role of the gut microbiome in the nutrition-health paradigm. Specifically, dietary carbohydrates have been shown to play an important role in human health through their interaction with the gut microbiome.[6, 7] Despite their importance and being one of the most abundant components in foods, their structures, abundances, and functions are still poorly characterized due to a general lack of appropriate analytical methods.[8] While the analysis of proteins and lipids have advanced greatly, the analysis of carbohydrates has been hindered by the complexity of carbohydrate structures. Current methods categorize carbohydrates into the broad and widespread classifications of sugars, starch, and soluble/insoluble fiber. Furthermore, the term "fiber" gives no monosaccharide or structural specificity. The total carbohydrates in most foods are measured indirectly by gravimetric mass

difference of other macronutrients and micronutrients.[9] Current analytical methods make it difficult to resolve the relationships between carbohydrates, the gut microbiome, and host health. There is thus a need for high throughput methods that is capable of characterizing carbohydrate structures and their microbiome interactions in large feeding studies.[10]

A major challenge in understanding the specific role of carbohydrates is the lack of knowledge of carbohydrate structures. Food carbohydrates are comprised of a diverse set of molecules ranging from free monosaccharides, disaccharides, oligosaccharides, and large polysaccharides. Additionally, each monosaccharide residue connects to another through numerous linkages (as many as 10 for each glycosidic linkage). Methods for oligosaccharide analysis using liquid chromatography-tandem mass spectrometry (LC-MS/MS) have been developed for structural elucidation, however the analyses remain very difficult and require a number of separation steps and structural elucidation techniques.[8] Furthermore, even the most fundamental information, the monosaccharide composition is not known in most foods. The lack of this basic structural information inhibits our understanding of the role of the most abundant material in our diet. It prevents effective design of important clinical trials that could elucidate the specific roles of specific carbohydrate structures in food.

In this research, a recently developed workflow utilizing a high-throughput UPLC-QqQ-MS method was employed to determine the monosaccharide compositions of over 800 food samples. Foods from diverse groups such as fruits, vegetables, fats, grains, dairy, beverages, and processed foods were subjected to monosaccharide analysis, and the resulting monosaccharide compositions were used to create a foundational collection of monosaccharide compositions or the glycan encyclopedia (Glycopedia). The method entails the absolute quantitation of 14

naturally occurring monosaccharides separated on a five-minute UPLC-QqQ MS analysis in a 96-well plate format. This platform and the resulting Glycopedia allows us to formulate feeding trials where the diets may be highly enriched for specific monosaccharide residues. Tailoring diets will enable future studies to better understand the role of food carbohydrates in shaping the gut microbiome in infants and adults. Furthermore, the presented findings will allow for dietary interventions that are more precisely formulated for modulating the gut microbiome and impacting human health.

MATERIALS AND METHODS

Selection of foods for inclusion in the Glycopedia. Foods were initially selected for the Glycopedia based on designing a feeding trial to selectively enrich beneficial gut microbiota in toddlers (12-36 months). The toddler foods selected for the Glycopedia include single foods recommended for toddlers according to the 2020-2025 Dietary Guidelines for Americans[11] which include a diverse group of vegetables that also contain appreciable amounts of arabinose such as dark and green vegetables; red and orange vegetables; beans, peas, and lentils; in addition to lower arabinose-containing vegetables, fruits and starches. In addition to analyzing single foods, the Glycopedia includes food mixtures and snacks that contain various levels of arabinose. Additional foods were then selected to cover foods that are commonly consumed by adults.

To determine foods commonly consumed by adults, three datasets were reviewed: (1) the Nutritional Phenotyping study (NutPheno),[12] (2) What We Eat in America (WWEIA) 2017-2018,[13] and (3) the Food and Nutrient Database for Dietary Studies Ingredients Database (FNDDS-Ing).[14] The NutPheno study was a cross-sectional study that included healthy male and

female adults, aged 18-66 y, living near Davis, CA. The NutPheno study included 393 adult

subjects who reported dietary intake with up to 4 days of 24-hour recalls using the Automated

Self-Administered 24-hour Dietary Assessment Tool (ASA24).[15] WWEIA is the dietary

component of the National Health and Nutrition Examination Survey (NHANES), a nationally-

representative cross sectional study and consists of two 24-hour dietary recalls. Both the

WWEIA dietary assessment and ASA24 use the Food and Nutrient Database for Dietary Studies

(FNDDS), and the food descriptions and numeric identifiers (Food Code) come from FNDDS.

The FNDDS-Ing database file describes the ingredients used to build FNDDS mixed foods

(recipes), and therefore lists ingredients that might otherwise not be reported through WWEIA or

ASA24 (in which subjects typically report final dishes instead of each individual ingredient in a

food).

In the NutPheno study, a total of 2435 unique foods (corresponding to 2435 unique

FNDDS Food Codes) were reported from a total of 1499 recalls. To identify candidate adult

foods to add to the Glycopedia from the NutPheno study, the frequency of each food reported in

NutPheno was counted, and the 200 most frequently reported foods was manually cross-matched

to the Glycopedia by searching the food description in the Glycopedia for the closest match. Of

the top 200 most frequently consumed foods, 135 did not have a matching Glycopedia food. A

second round of manual curation was conducted on these 135 NutPheno foods to identify

candidate foods to add to the Glycopedia (e.g. would likely contribute to dietary glycan

consumption and/or are typically consumed in large quantities or very frequently, n = 59).

The same process described above for NutPheno foods was used for FNDDS-Ing and

WWEIA. A total of 2744 unique ingredients were identified in FNDDS-Ing. The frequency of an

ingredient corresponds to the total number of times the ingredient is used in FNDDS recipes. Of

the top 200 most frequently reported ingredients, 137 did not have matches to the Glycopedia, 79 of which were considered as candidates to add to the Glycopedia. A total of 7083 foods were reported in WWEIA. Of the top 200 most frequently consumed foods, 130 had no Glycopedia match, and 49 were considered as candidates to add to the Glycopedia.

The candidate foods from NutPheno, FNDDS-Ing, and WWEIA were compared and discussed by the study team. Foods that were frequently consumed and with potentially high glycan content were prioritized and selected to add to the Glycopedia, and candidate foods that were consumed in relatively low amounts or could be approximated using other Glycopedia foods were given low priority.

All foods and food products were purchased from local markets (Davis and Sacramento, CA) including Safeway, Trader Joe's, Davis Food Co-op, Whole Foods, Nugget Markets, Target, and online (Amazon). Trifluoroacetic acid (TFA) (HPLC grade), 3-methyl-1-phenyl-2-pyrazoline-5-one (PMP), chloroform (HPLC grade), ammonium hydroxide solution (NH_4OH) (28-30%), ammonium acetate, sodium acetate, glacial acetic acid, methanol (HPLC grade), D-fructose, D-mannose, D-allose, D-glucose, D-galactose, L-rhamnose, L-fucose, D-ribose, D-xylose, L-arabinose, N-acetyl-D-glucosamine (GlcNAc), N-acetyl-D-galactosamine (GalNAc), D-glucuronic acid (GlcA), and D-galacturonic acid (GalA) were purchased from Sigma-Aldrich (St. Louis, MO). 96-well Nunc plates were purchased from Thermo Scientific. Arabinoxylan was purchased from Megazyme (Bray, Ireland). Viscozyme® was provided by Novozyme (Davis, CA). Acetonitrile (ACN) (HPLC grade) was purchased from Honeywell (Muskegon, MI). Nanopure water was used for all experiments.

Preparation of food samples. A total of 800 foods including fresh, frozen, commercial, and processed were purchased from local grocery stores in Davis, CA. Each food was documented with detailed descriptions prior to the sample preparation. Most foods were aliquoted raw for analysis. For some raw and packed foods, samples were cooked, baked, or steamed as indicated in the package for cooking instructions. Foods were lyophilized to complete dryness and the moisture content was obtained. Samples underwent a dry bead blast or mortar and pestle to homogenize the samples. A 10 mg aliquot of dried food sample was weighed into a 1.5 mL screw cap Eppendorf tube and was reconstituted with water to make a stock solution of 10 mg/mL. The stock solution then underwent a bullet blending procedure follow by heat treatment (1 h at 100°C) and another round of bullet blending prior to monosaccharide analysis.

Monosaccharide analysis of food samples. The monosaccharide analysis of foods was adapted from Xu et al.[16] and Amicucci et al.[17] with the following modifications. A 10 uL aliquot from the homogenized stock solution was subjected to incubation with Viscozyme (Novozyme, Davis, CA) treatment at 50°C for 1 hr in 390 µL of 25mM acetate buffer (pH 5). A 100 µL aliquot from the enzyme digest was subjected to hard acid hydrolysis with 4 M TFA for 1 h at 121°C and quenched with 855 µL of ice-cold water. A pool of monosaccharide standards consisting of D-fructose, D-mannose, D-allose, D-glucose, D-galactose, L-rhamnose, L-fucose, D-ribose, D-xylose, L-arabinose, N-acetyl-D-glucosamine (GlcNAc), N-acetyl-D-galactosamine (GalNAc), D-glucuronic acid (GlcA), and D-galacturonic acid (GalA) were to generate a calibration curve and were prepared in water ranging in concentration from 0.001 to 100 µg/mL. The released monosaccharides in samples and standards were then derivatized with 0.2 M PMP solution in methanol and 28% NH$_4$OH at 70°C for 30 min. Samples were then dried to completeness by

vacuum centrifugation. The excess PMP was removed by a chloroform extraction. A 1 μL

aliquot of the derivatized monosaccharides were subjected to UPLC-QqQ MS analysis.

Mass spectrometry instrumental analysis. Derivatized glycosides were separated on an

Agilent Poroshell HPH-C18 column (2.1 x 50 mm, 1.9 um) and guard using an Agilent 1290

Infinity II UPLC system. A constant flow rate of 1.050 mL/min was employed on a 2 min

isocratic elution at 12% solvent B followed by a 1.6 min flush at 99% solvent B and 0.79 min

equilibration for a total run time of 4.6 min for the separation of compounds. Solvent A

consisted of 25mM ammonium acetate in 5% acetonitrile. Solvent B consisted of 95%

acetonitrile in water. The separated glycosides were then detected on an Agilent 6495B triple-

quadrupole mass spectrometer (QqQ-MS) operated in positive ion mode using dynamic multiple

reaction monitoring (dMRM).

Data analysis. Raw LC-MS files were analyzed using Agilent MassHunter Quantitative

Analysis software (Version B 08.00). Chromatographic peaks were manually integrated and

matched with standards. Monosaccharides were quantified by external calibration curve fitted

with linear regression. Clustering analysis based on monosaccharide profiles were done with R

using circlize library (v 0.4.13). Dendrograms and heatmaps in **Figure 2.8a** were also generated

using circlize. Enrichment of food groups in each cluster was determined using hypergeometric

test and statistical significance was assigned based on FDR-adjusted p-values.

Assigning food groups to Glycopedia foods. The Glycopedia food groups are adapted from the

FNDDS food groups that are defined by the first two digits of the FNDDS Food Code:[18] (1) Milk

and Milk Products, (2) Meat, Poultry, Fish, and Mixtures, (3) Eggs, (4) Beans, Peas, Other

Legumes, Nuts, and Seeds, (5) Grain Products, (6) Fruits, (7) Vegetables, (8) Fats, Oils, and

Salad Dressings, and (9) Sugars, Sweets and Beverages. Food groups were assigned based on a Glycopedia food's first ingredient. For example, both mango juice and fresh yellow mango are Fruits, and orange preserves (first ingredient is sugar) is in Sugars, Sweets, and Beverages. Two different study personnel (Y.B. and S.B.) assigned food groups to all foods. A third person (E.C.) compared the assignments. The foods with different assignments were re-evaluated by Y.B. and S.B. A final comparison and check of the food groups was done by E.C.

RESULTS

We employed a recently developed rapid-throughput LC-MS based method to determine and quantitate the total monosaccharide composition of over 800 foods. The resulting compositions provided insight into both the quantities and structural characteristics of the sugars, oligosaccharides, and polysaccharides that are present in the complementary foods commonly consumed by toddlers and older subjects in the US population. The resulting glycan composition of foods were used for cluster analysis and grouping of foods based on the monosaccharide profile. For the first time, the Glycopedia can be used to create personalized diets based on the absolute monosaccharide amounts in meals for the purpose of altering the gut microbiome.

Monosaccharide compositional analysis in foods

Foods purchased in local markets were documented with detailed descriptions and were processed using a sample preparation procedure that entailed cooking (where applicable), lyophilization, and dry homogenization to ensure homogenous sampling. Moisture content was determined at the lyophilization step. A 10 mg aliquot was then taken to prepare a stock suspension in water, which was further subjected to homogenization via heating and bullet bead blending. Liquid aliquots were then further subjected to a rapid, optimized monosaccharide

compositional analysis in 96-well plate format. This consisted of an enzyme digestion targeting

polysaccharides containing galacturonic acid (GalA) and subsequent acid hydrolysis to

breakdown oligo- and polysaccharides into monosaccharides as quantitatively as possible. The

liberated monosaccharides were then derivatized to make them more amenable to reversed-phase

chromatography and mass spectral analysis. The absolute monosaccharide composition of each

food were then be determined in a 5 minute UPLC-QqQ-MS analysis.
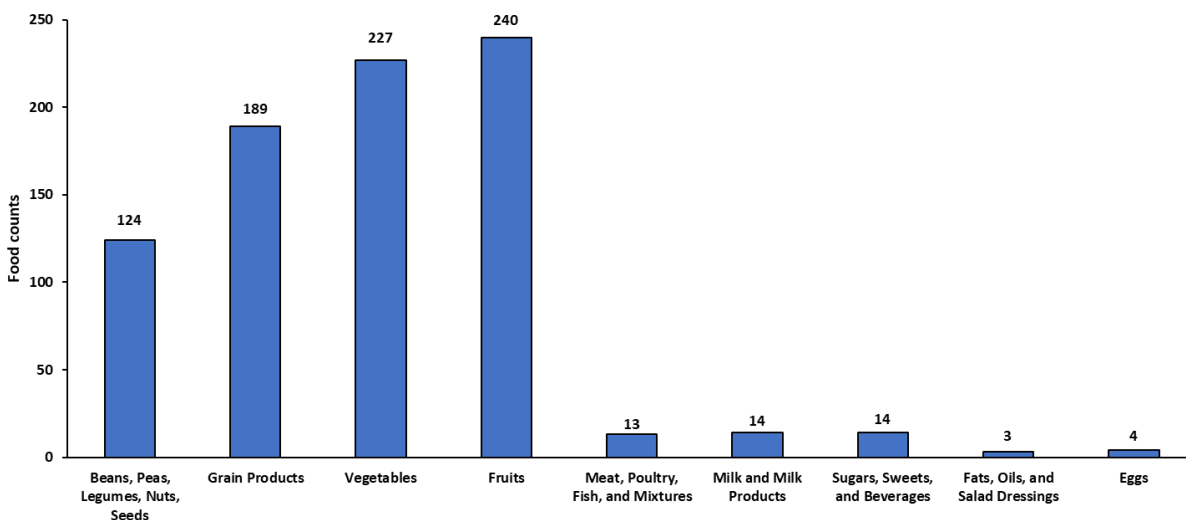
Figure 2.1. Number of foods assigned to each of nine total food groups. The number of foods

analyzed in each food group is noted above each bar graph. Those groups containing plant-based

foods contributed the largest number while animal-based and low carbohydrate groups

contributed the least.

Foods were assigned to one of nine food groups as detailed in the Methods section. The

number of food samples within each food group are summarized in **Figure 2.1**. The largest

groups were fruits, vegetables, grain products, and beans/peas/legumes/nuts/seeds, while the

groups with the smallest representations were those containing animal-based products, fats/oils/salad dressings, and sugars/sweets/beverages. **Figure 2.2a-j** illustrates the average monosaccharide profile obtained for each food group. Of the 14 monosaccharides monitored, glucose and fructose accounted for the majority of the monosaccharides measured. They were present in their free forms, as sucrose, and as starch. Other common and abundant monosaccharides included xylose, arabinose, galactose, and GalA, which correspond to cell wall polysaccharides such as arabinoxylan and pectins.[19, 20] Xylose was most abundant in grains (**Figure 2.2e**) along with arabinose due to the presence of arabinoxylans in the cell walls of grains, while beans, fruits, and vegetables (**Figure 2.2a, d, i**) tended to contain more GalA and rhamnose as part of pectins.[19, 20] Grain products had the highest overall measured carbohydrates by fresh weight (**Figure 2.2e**) due to high starch and low moisture content. This was followed by beans, peas, and legumes, fruits, and vegetables (**Figure 2.2a**). Eggs and fats/oils/salad dressings (**Figure 2.2b and 1c**, respectively) were found to contain the lowest carbohydrate content. The meat/poultry/fish group contained significant amounts of glucose likely from free glucose and glycogen (**Figure 2.2f**) contrary to what the carbohydrate content listed on their nutrition labels often suggest.
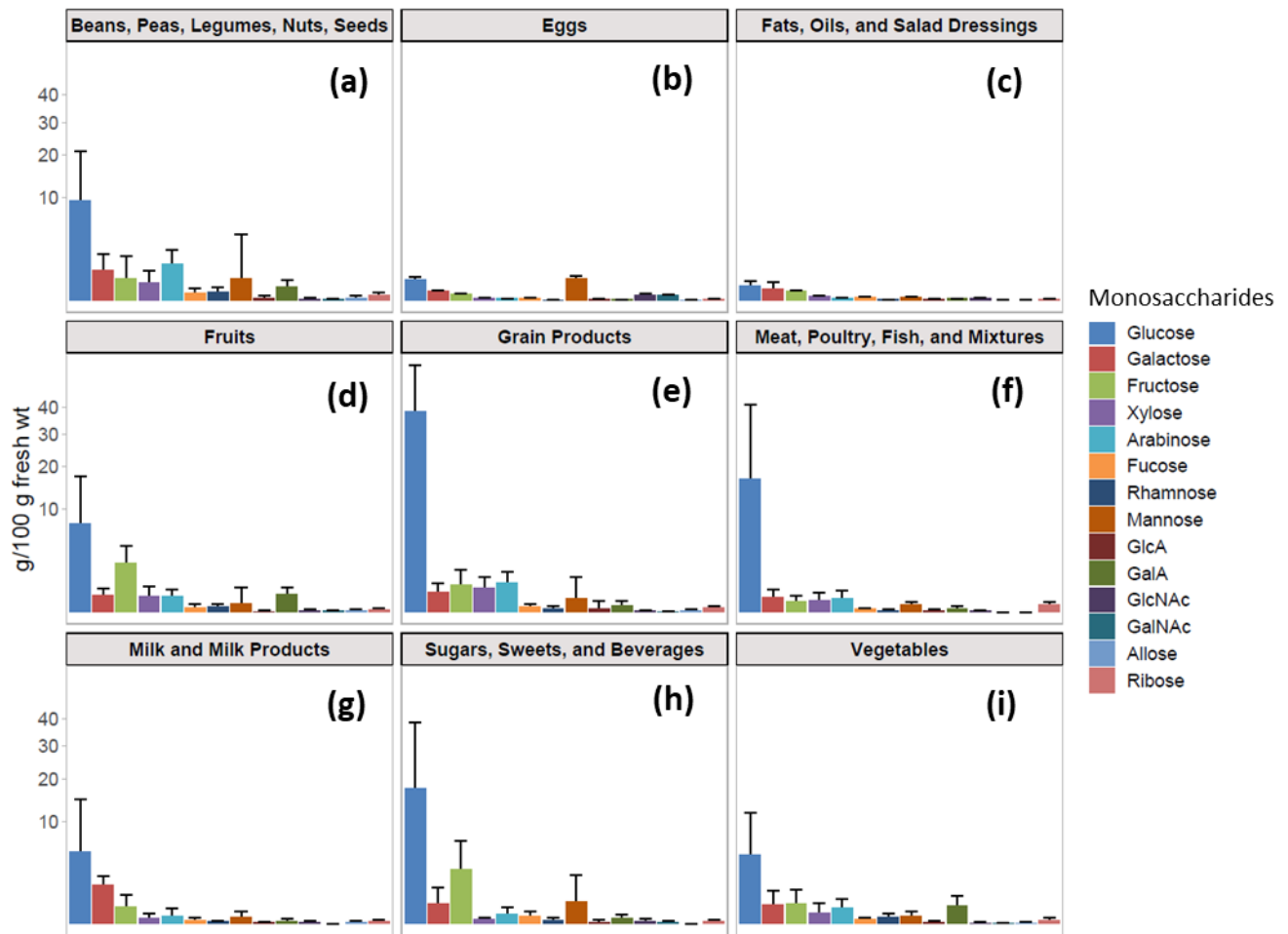
Figure 2.2a-j. Average monosaccharide compositions of all nine food groups. The y-axis follows a square root scale. Error bars represent the standard deviation.

Each plant-based food group had markedly different monosaccharide compositions. **Figure 2.3a-d** depicts the monosaccharide compositions of 20 representative foods from each plant-based food group. Fruits (**Figure 2.3a**) exhibited diverse monosaccharide compositions and tended to contain significantly more fructose than other groups due to a higher sugar content. Grain products (**Figure 2.3b**) exhibited the highest glucose content from starch, but also contained xylose, arabinose, and galactose. Aside from high-starch vegetables such as potatoes and corn, vegetables (**Figure 2.3c**) had diverse monosaccharide profiles similar to that of fruits

consisting of glucose, fructose, galactose, xylose, arabinose, GalA, and mannose while

containing markedly less fructose compared to fruits. Beans, peas, legumes, and nuts (**Figure**

2.3d) were all similar in that they tended to contain relatively high amounts of arabinose.

However, beans and peas had larger amounts of glucose than nuts due to a higher starch content.
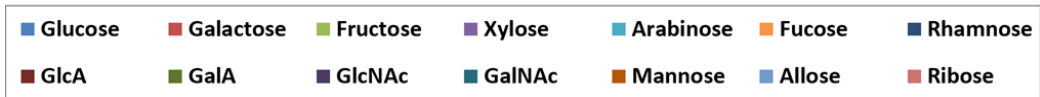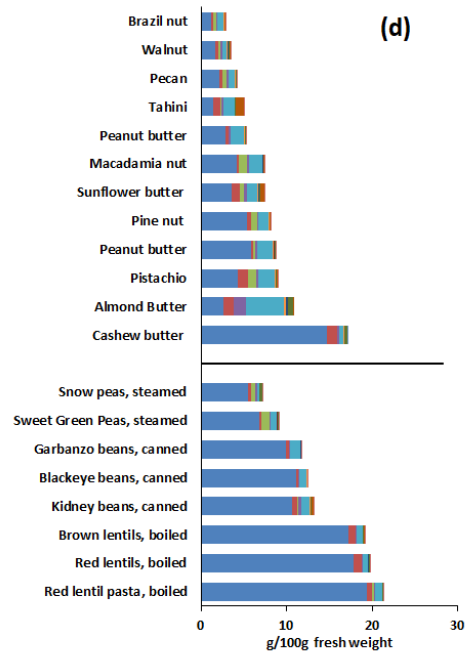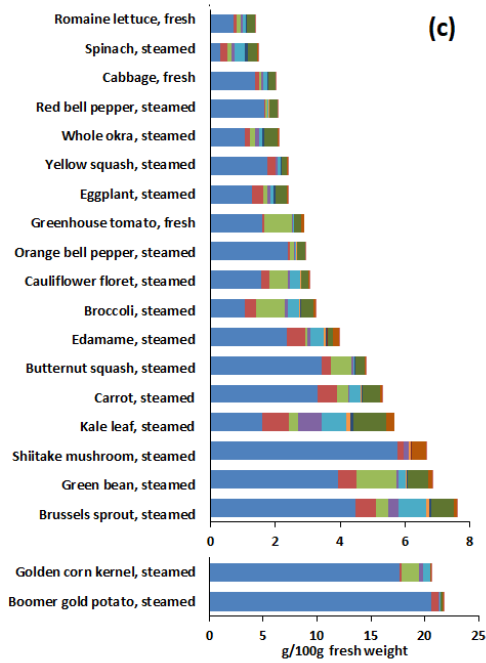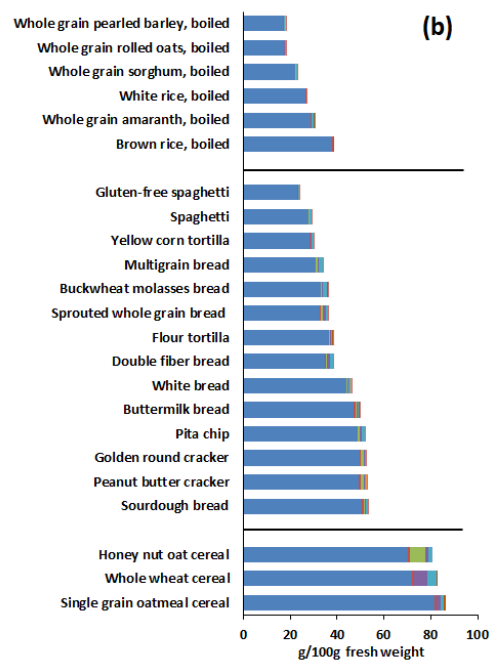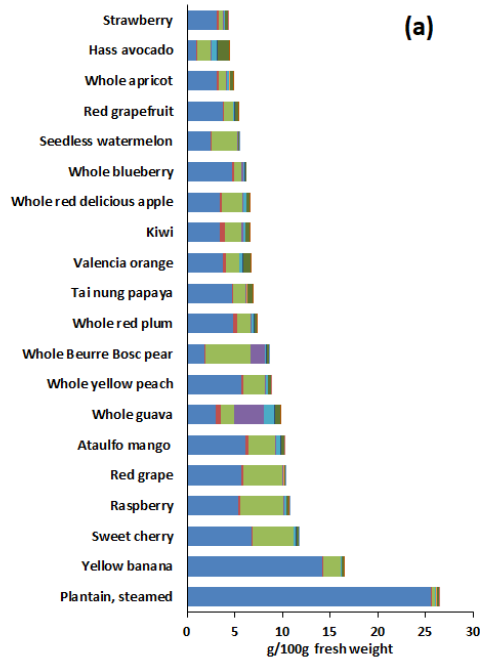
Figure 2.3a-d. Monosaccharide compositions of selected representative foods from each plant-based food group for fruits (a), grains products (b), vegetables (c), beans, peas, legumes, and nuts (d).

Foods within each food group also exhibited varying compositions. Solanaceous foods (or nightshades) like tomatoes, eggplant, and bell peppers (**Figure 2.3c**) contained very little arabinose and non-glucose glycans while members of the Brassicaceae family like brussels sprouts, broccoli, kale, and cauliflower (**Figure 2.3c**) exhibited significantly larger quantities of arabinose and other glycans such as GalA, galactose, rhamnose, and fucose. Similar differences were observed in the other food groups. In nuts, almonds contained the largest amount of arabinose while tahini (made from sesame seeds) contained large amounts of mannose (**Figure 2.3d**). In fruits, pears and guava tended to contain more xylose than other fruits while berries were very low in non-glucose and non-fructose monosaccharides (**Figure 2.3a**). In grains, "white" products like white bread, flour tortillas, and white rice tended to contain less non-glucose monosaccharides than their whole grain counterparts like whole-grain bread, grains, and brown rice (**Figure 2.3b**).

Monosaccharide composition of diets

The results from the Glycopedia were used to quantitate carbohydrates in a standard diet. According to the USDA Dietary Guidelines for Americans 2020-2025 and USDA MyPlate, it is recommended for adults to consume 2 cups of fruits, 2.5 cups of vegetables, 6 ounces of grains, 5.5 ounces of protein, and 3 cups of dairy in a day.[11] These recommendations are based on consuming 2,000 calories per day and have different food groups compared to the food groups described in this work. To generate a relative chart of each food group (**Figure 2.4**), the

recommended servings in an example meal was converted from cups and ounces to grams. The ingredients for the example dinner meal included 4 ounces of chicken breast, 0.5 cups of broccoli, 0.33 cups of carrots, 0.33 cups of summer squash, 0.75 cups of pasta, 1 tablespoon of oil, 1 cup of a navel orange, and 1 cup of milk. Based on the USDA Dietary Guidelines for Americans 2020-2025, the recommended food groups relative composition for the example meal yielded 30 % for Vegetables, 26 % for Fruits, 19 % for Dairy, 13 % for Grains, and 12 % for Proteins.
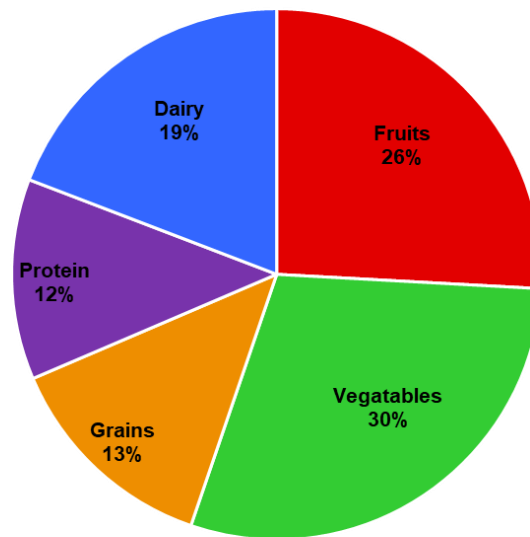
Figure 2.4. Mass percentages of food groups used to generate an example meal based on recommendations in the USDA Dietary Guidelines for Americans 2020-2025.
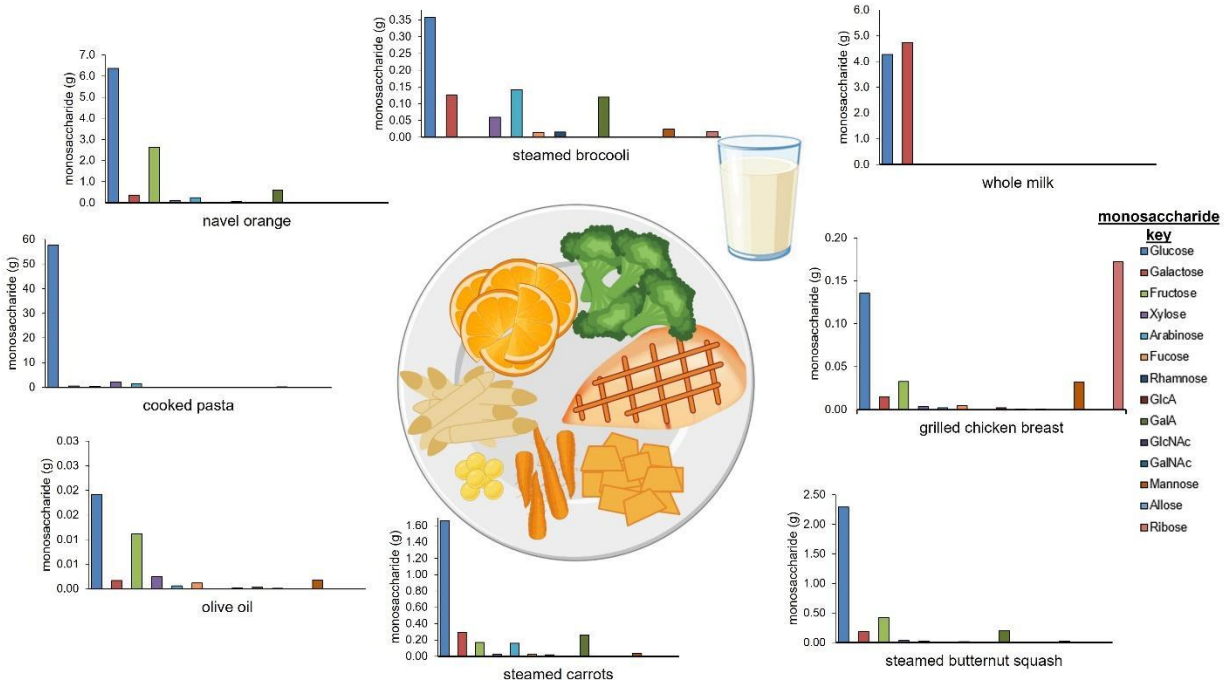
Figure 2.5. Example meal with quantitative monosaccharide bar graphs of each ingredient. The serving amounts are based on the USDA Dietary Guidelines for Americans 2020-2025.

The total dietary carbohydrate content in the example meal was determined using values from the Glycopedia. Additionally, the monosaccharide concentrations and composition of each ingredient in the meal were determined (**Figure 2.5**). The calculated total carbohydrate content in the entire meal was 89.09 g (**Table 2.1**). The cooked penne pasta, navel orange, and glass of whole milk resulted in the highest total carbohydrate amounts (per ingredient and serving) with values of 62.4 g, 10.4 g, 9.1 g, respectively. As expected, olive oil, and grilled chicken breast had minimal carbohydrates (per ingredient and serving) with values 0.0 g, and 0.4 g, respectively. The cooked penne pasta had less relative monosaccharide diversity with glucose from starch as the most abundant. On the other hand, steamed broccoli, steamed carrots, navel orange and steamed butternut squash had the most (non-glucose) monosaccharide diversity with higher amounts of galactose, fructose, xylose, arabinose, and galacturonic acid present. The whole milk

`contained glucose (4.27 g) and galactose (4.74 g) per 1 cup of whole milk and matched the known composition of lactose.

Table 2.1. The absolute monosaccharide composition and amounts in an example dinner meal.

food	serving amount	amount (grams)	moisture (%)	monosaccharide (g)														
				Glc	Gal	Fruc	Xyl	Ara	Fuc	Rhm	GlcA	GalA	GlcNAc	GalNAc	Man	All	Rib	total
grilled chicken breast	4 oz	113.4	62	0.14	0.01	0.03	0	0	0	0	0	0	0	0	0.03	0	0.17	0.4
steamed broccoli	0.5 cups	38	89.6	0.36	0.13	0	0.06	0.14	0.01	0.02	0	0.12	0	0	0.02	0	0.02	0.9
steamed carrots	0.33 cups	50	87.1	1.66	0.29	0.17	0.03	0.16	0.02	0.02	0	0.26	0	0	0.04	0	0	2.7
steamed butternut squash	0.33 cups	66.7	89.6	2.29	0.19	0.42	0.03	0.03	0.01	0.01	0	0.2	0	0	0.02	0	0	3.2
cooked pasta	0.75 cups	150	54.7	57.6	0.51	0.45	2.1	1.51	0	0.02	0.01	0.03	0	0	0.15	0	0.03	62.4
olive oil	1 Tbsp	14.3	0.8	0.02	0	0.01	0	0	0	0	0	0	0	0	0	0	0	0
navel orange	1 medium orange	165	87.3	6.37	0.37	2.64	0.12	0.24	0.04	0.06	0	0.6	0	0	0	0	0	10.4
whole milk	1 cup of milk	245	89	4.27	4.74	0	0	0.01	0	0	0	0	0	0	0.03	0	0.01	9.1
total	N/A	N/A	N/A	72.7	6.25	3.72	2.35	2.09	0.09	0.13	0.01	1.21	0	0	0.3	0	0.25	89.09

In addition to determining the total carbohydrate content in each ingredient in a meal, the database was used to determine the total amount of each monosaccharide by adding the total monosaccharides from each ingredient. In the exemplified meal above, the glucose was the most abundant monosaccharide with a total of 72.70 g. The next most abundant monosaccharides were galactose and fructose with a total of 6.25 g and 3.72 g, respectively. Xylose (2.35 g) and arabinose (2.09 g) were similar in abundance, while fucose (0.09 g), rhamnose (0.13 g), galacturonic acid (1.21 g), mannose (0.30 g), ribose (0.25 g) were present in smaller amounts. With the Glycopedia, the monosaccharide profile of a meal can be altered by simply swapping an ingredient from the same food group with a higher concentration of the desired monosaccharide.

For example, if more arabinose is desired the navel orange (0.15 g/100g arabinose) from the meal above can be exchanged for a Bartlett pear (0.37 g/100g arabinose).

Personalized nutrition based on specific monosaccharide abundances

The Glycopedia can be used not only to compare foods but to create personalized meals rich in specific monosaccharides and, by extension, fibers for altering and modulating the gut microbiome or other nutritional studies. Arabinose is a prime target for this purpose as it is found commonly only in plants, is not digested or absorbed endogenously, and has been shown to play an important role in shaping the gut microbiome.[21] Arabinose is not abundant in foods as a free monomer, rather it is a part of ubiquitous cell wall polysaccharides such as arabinoxylan in grains and pectins in fruits and vegetables.[19, 22] While this method does not differentiate from which polymer the arabinose originates, arabinose can nonetheless be quantitated to identify foods to maximize dietary this saccharide. **Figure 2.6** provides the broad arabinose content of the individual food groups. The highest average arabinose content was observed in Food Group 1 (1.24g/100g fresh weight) followed by groups 2, 3, and 4 (0.8, 0.27, and 0.24g/100g fresh weight). In general, the highest arabinose concentrations were found in plant-based foods such as legumes, grains, vegetables, and fruits. However, the range of arabinose in each plant-based food group was large and depended on the specific food and moisture content. For example, pear cultivars tended to have more arabinose than apple cultivars. Relatively dry foods like cereals, nut butters, and dehydrated legume, vegetable, and fruits products consistently displayed the highest arabinose concentrations and total measured carbohydrate in each group.
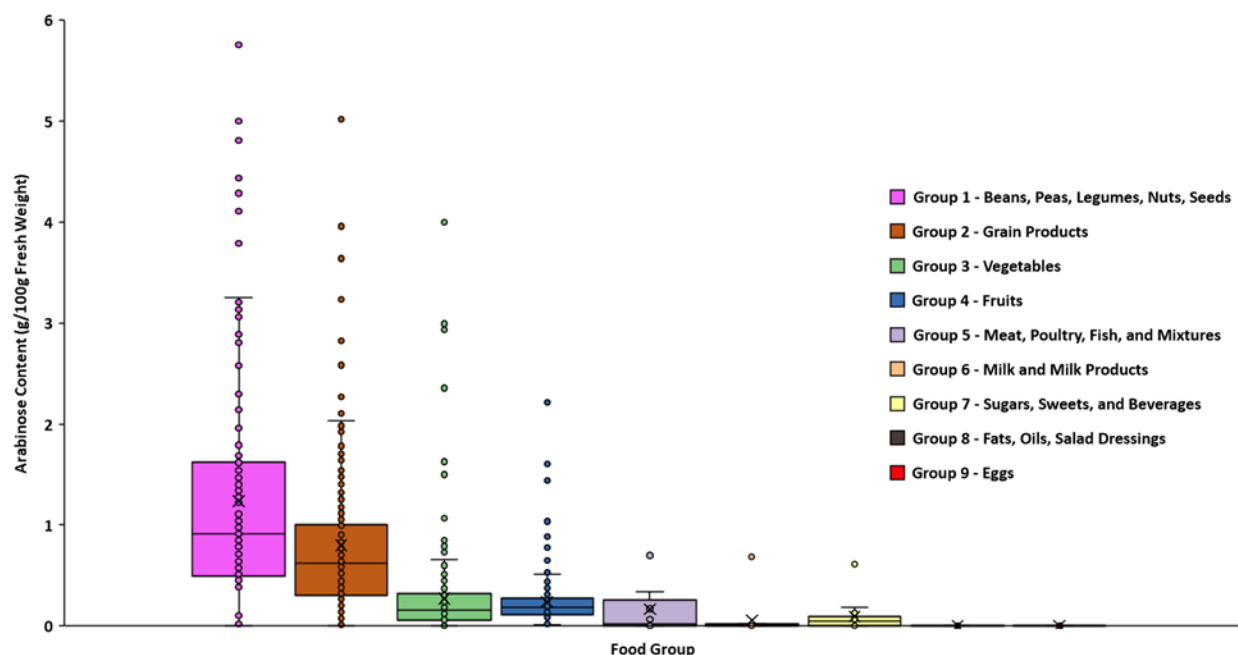
Figure 2.6. Average arabinose abundances in all food groups. Food group 1 yielded the highest arabinose amounts and group 9 yielded the least, respectively.

Processed foods – monosaccharide composition in commercial complementary foods

To investigate the carbohydrate content in processed foods or foods containing multiple ingredients such as commercial complementary foods, we compared the levels of arabinose in 23 products in a subset of a store name brand. In processed foods, multiple ingredients were used to make the final product where the monosaccharide abundances of ingredients vary. Because the ingredients mostly contained raw food ingredients, the arabinose concentration (by fresh weight) for that whole food was used to generate the heat map (**Figure 2.7**) in commercial complementary foods. For example, the raw ingredients for "Happy Tot Super Foods: pears, mangoes, spinach, super chia" included raw pears, mangoes, and spinach and were found to contain 0.27, 0.67, and 0.17 g of arabinose/100 g of fresh weight, respectively. Among this product, the arabinose content in mango (0.67 g/100 g fresh weight) was highest from all

ingredients, while the total arabinose content of the complementary food product had lower

amounts (0.47 g/100 g fresh weight). For the rest of the complementary food for babies and

toddlers had a varying range from 0.097 to 0.77 g/100 g fresh weight.

Product	Ingredient 1	Ingredient 2	Ingredient 3	Ingredient 4	Ingredient 5
bananas, sweet potatoes, & papayas	banana	sweet potato	papaya	lemon juice	vitamin C
pears, squash, & blackberries	pear	butternut squash	blackberry	vitamin C	lemon juice
apples, guavas, & beets	apple	guava	beet	vitamin C	lemon juice
apples, pumpkin, & carrots	apple	pumpkin	carrot	vitamin C	lemon juice
apples, blueberries, & oats	apple	blueberry	oats	vitamin C	lemon juice
bananas, raspberries, & oats	banana	raspberry	oats	vitamin C	lemon juice
apples, kale, & avocados	apple	kale	avocado	vitamin C	lemon juice
pears, kale, & spinach	pear	kale	spinach	vitamin C	lemon juice
bananas, pineapple, avocado, & granola	banana	pineapple	oats	avocado	quinoa
zucchini, apples, peas, quinoa, & basil	zucchini	apple	peas	quinoa	basil
bananas, plums, & granola	banana	plum	oats	amaranth	vanilla extract
purple carrots, bananas, avocados, & quinoa	purple carrot	banana	avocado	quinoa	vitamin C
apples, sweet potatoes, & granola	apple	sweet potato	oats	quinoa	vanilla extract
oatmeal, sprouted quinoa, bananas, strawberries	oats	apple	blueberry	quinoa	chia seed
oatmeal, sprouted quinoa, apples, blueberries	oats	banana	apple	strawberry	quinoa
squash ravioli with squash, pumpkin, & sage sauce	tomato	sweet potato	pear	spinach	tomato
cheese & spinach ravioli with marinara sauce	butternut squash	pumpkin	apple	wheat flour	corn starch
fiber & protein- pears, raspberries, carrots, butternutsquash	pear	raspberry	carrot	butternut squash	pea protein
super foods- apples, sweet potato, carrots, cinnamon, super chia	apple	sweet potato	carrot	chia seed	cinnamon
super foods- pears, beets, blueberries, super chia	pear	beet	blueberry	chia seed	lemon juice
super foods- pears, mangos, spinach, super chia	pear	mango	spinach	chia seed	lemon juice
fiber & protein- pears, kiwi, & kale	pear	kiwi	kale	pea protein	chia seed
super foods- pears, green beens, peas, super chia	pear	green beans	peas	chia seed	lemon juice

Figure 2.7. Heat map of 23 commercial complementary foods from Happy Family brand. The

bar graphs on the left represents the total amount of arabinose found in the complementary food

product along with the list of corresponding ingredients on the right with arabinose content in the

ingredient whole food for babies (a) and toddlers (b).

The first whole food ingredient in processed foods contributes greatly to the

monosaccharide composition. For example, when bananas were the first ingredients in the

"Happy Family" infant products, the total arabinose content was low (less than 0.2 g of

arabinose/100 g fresh weight). On the other hand, when pears were the first ingredients, the total

arabinose had greater than 0.2 g of arabinose/100 g fresh weight with exception of the "Super

Foods: pears, green beans, peas, super chia" product. The cheese & spinach ravioli with marinara

sauce meal in the "Happy Family" toddler product yielded the highest arabinose content, likely due to the minimal moisture content.

Monosaccharide and Clustering Analyses

To visualize common features and differences among individual monosaccharide compositions irrespective of food group, we performed an unsupervised hierarchical clustering analysis. The average monosaccharide compositions of each cluster are depicted in **Figure 2.8b-k**. A total of 10 clusters were chosen to divide the 828 foods into clusters based on their total monosaccharide compositions. Food belonging to the same groups (fruits, vegetables, grain products, etc.) largely clustered together and were defined by their monosaccharide compositions. Cluster 1 was the largest, comprising of over half of the total foods surveyed. This cluster was significantly enriched in fruits and vegetables, but also contained legumes, plant-based milks, and few members of the sugars, sweets, and beverages food group. The average monosaccharide composition for this cluster (**Figure 2.8b**) was most dissimilar to clusters 2-9 (**Figure 2.8c-j**, respectively) and reflected the fruit and vegetable food groups. Specifically, cluster 1 contained significantly lower amounts of glucose and a larger overall diversity than other clusters. Cluster 2 was significantly enriched in grain products which were largely cooked whole grains such as oats, barley, millet, quinoa, and rice. Additionally, cluster 2 contained bean-based pastas and dried fruits. Cluster 3 was significantly enriched in foods from the beans, peas, legumes, nuts, and seeds group, but like Cluster 1 also contained fruits, vegetables, and grain products. Clusters 4-9 were all significantly enriched in grain products. However, each cluster differed in their average monosaccharide compositions, namely their glucose content. Cluster 4 contained the least average glucose, the highest average fructose, and grouped closer to cluster 2 which contained many cooked whole grains and dried fruits. Clusters 5-9 contained the

highest average glucose and total carbohydrate and consisted mostly of dried grain products such as cereals and snacks. Cluster 10 contained only coconut flour due to its unique composition of mostly mannose.
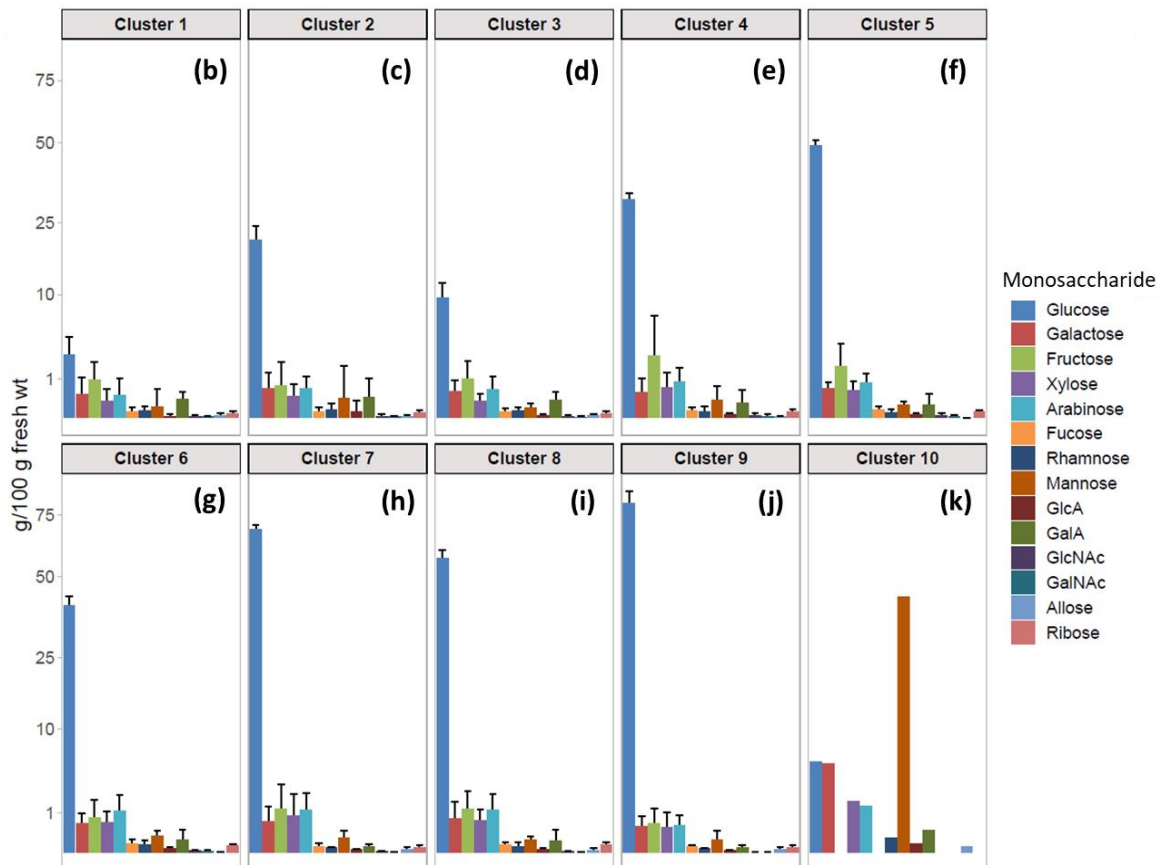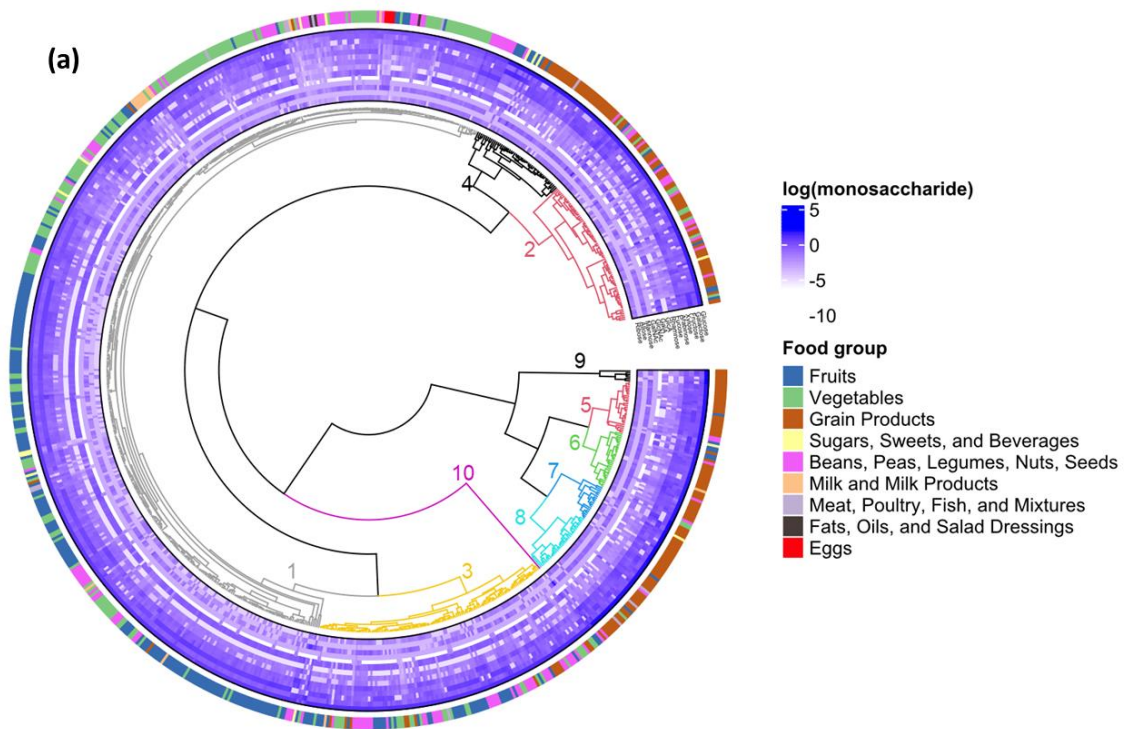
Figure 2.8. (a) Hierarchical cluster analysis of all 828 foods based on their absolute monosaccharide compositions, (b-k) Average monosaccharide composition of each cluster. The y-axis follows a square root scale. Error bars represent the standard deviation.

DISCUSSION

The Glycopedia revealed the most abundant monosaccharide in foods was primarily glucose from simple sugars such as sucrose and starch polysaccharides. Plants have evolved to pack seeds, grains, tubers, and fruits with glucose, while other components of the plant have significantly less glucose, but larger amounts of other monosaccharide residues such as arabinose, xylose, ribose, and mannose. From an evolutionary and agricultural perspective, humans have historically used innovative strategies to cultivate sugar- and starch-dense foods and parts of foods as a source of energy. While these energy-rich foods were once a necessity for survival, increasingly sedentary lifestyles, and overconsumption of highly processed versions of these foods has contributed to a variety of metabolic disorders such as obesity, type 2 diabetes, and heart disease particularly in Western populations. The Glycopedia provides information not only on digestible glucose content, but also on non-glucose content corresponding to various dietary fiber structures. This information can be used to inform dietary choices to alleviate these metabolic disorders by reducing starch and sugar consumption and increasing the consumption of specific fiber types to shape the gut microbiome in a targeted manner. In early life for infants, HMOs play a large role in shaping the infants gut microbiome. The next source of carbohydrates in an infant's diet are from plant-based foods containing dietary fiber. Toddlerhood (12 to 36 months) is a window in life when the gut microbiome is unstable[23] and provides an opportunity to shape the gut microbiome with selective dietary plant oligosaccharides (e.g., prebiotics). Specifically, bifidobacterial species that utilize plant-derived carbohydrates are purported to

confer several beneficial properties on their host such as eliciting immune-modulation, enhancing gut barrier function[24] and thwarting the colonization of pathogenic microbiota.[25] The sequence of the bifidobacterial genome reveals extensive glycosyl hydrolase activities, of which several were predicted to be involved in arabinose oligosaccharide utilization.[26] Furthermore, comparative genomics of 40 bifidobacterial species reported a highly diverse gene repertoire involved in the catabolism of arabino-oligosaccharides even in phylogenetically close species.[27] Currently, supplementing infants' diets with commercially-available complementary foods is widely accustomed and practiced. The commercial complementary foods analyzed in this work, provided the monosaccharide compositional information that can be used to modulate the microbiome based on the monosaccharide diversity and abundances. This information can thereby influence the infant's health and development.

Our research also demonstrated that monosaccharide compositions can vary within food groups with several implications for nutrition research. It will be necessary for nutrition studies to resolve their dietary data at the individual food level, rather than summarizing servings at the food group level, if the intent is to study food-microbiome structure relationships. Mixed meals will need to be resolved at the ingredient level. Finally, the database will eventually need to be expanded to incorporate the full variety of plants consumed. This is both a blessing and a curse; it means more work, but it also means more opportunity to alter human health via specific breeding or selection of specific existing varietals.

While the human genome contains just 17 enzymes for the saccharolytic dissection of dietary carbohydrates, the microbial genomes of the gastrointestinal tract include thousands of enzymes,[28] underscoring the importance of elucidating the structure of dietary carbohydrates in order to understand the relationship between diet and microbiome. Further, numerous lines of

evidence demonstrate that specific oligosaccharides and polysaccharides have direct effects on human cells, even in the absence of microbes, influencing intestinal barrier function and inflammation *in vitro*.[29-34] Thus, the carbohydrate aspect of diet potentially has far-reaching effects that we can now begin to investigate in the context of whole foods, rather than in the context of isolated polysaccharides such as inulin.

The broad epidemiologic importance of carbohydrates in the diet is well established. Their role in caloric transfer is critical to human health, but this simple view belies their important intrinsic biological activities. Increased consumption of carbohydrates that resist digestion by the host, typically termed "dietary fiber," has been associated with a reduced risk of obesity, type 2 diabetes, certain gastrointestinal disorders, and coronary heart disease.[35] Even monosaccharides, the smallest carbohydrate unit, have their own inherent activities.[36] More recently, the ability of carbohydrates to modulate the gut microbiome has become of considerable interest.[10, 37] While carbohydrates in food are undisputedly a necessary part of any healthy diet, the relative amounts, types of carbohydrates, and whether some foods can be called carbohydrates at all are the subject of considerable and even broad disagreements.[38] The conflicts regarding carbohydrates stem from our general ignorance of their chemical structures.

Current methods in dietary carbohydrate analysis are limited to quantifying sugars, starch, and fiber. Within the definition of fiber is an immense amount of structural complexity that can alter the gut microbiome and affect host health. Dietary recommendations emphasize the importance of consuming fiber. However, the term "fiber", including dietary and indigestible, makes no distinction of the monosaccharide composition nor the primary structure of the molecule. The reality is that food glycans are a very large number of compounds, each with their unique structural variations and potentially specific activities. Thus, the advice "eat more fiber,"

is not meaningful because fiber from two different sources can have completely different monosaccharide compositions, glycosidic bond linkages, degree of polymerization, and in turn, biological functions. The analytical methods used to measure carbohydrates must be updated to match the evolving throughput and coverage of sequencing and metabolomic analyses. To address this need, we developed and utilized a rapid-throughput, LC-MS based monosaccharide analysis to determine the total monosaccharide composition of 800 foods to create a food "Glycopedia" which will inform future feeding studies in infants transitioning to complementary diets, toddlers and adults. The total monosaccharide composition and quantitation provides more useful information on dietary carbohydrates than traditional gravimetric methods especially in the context of the gut microbiome and infant nutrition. This comes with greatly increased sample throughput making the construction of large food glycan libraries possible.

Even within the monosaccharide compositions determined in the Glycopedia, there lies an enormous amount of structural diversity as the polysaccharide and glycosidic linkage level information are not captured. Additionally, the methods utilized here did not employ sample preparation steps to separate free sugars and oligosaccharides from polysaccharides. Thus, for example, free fructose and glucose are not differentiated from inulin or starch, respectively. Future method development will utilize a rapid-throughput analytical workflow in which free saccharides and polysaccharides are identified and quantitated separately for a comprehensive and high-resolution picture of food carbohydrates.

CONCLUSION

A novel rapid throughput UPLC-QqQ-MS workflow was developed to determine the absolute quantitation of 14 monosaccharides in 800 foods. Foods from diverse groups such as fruits, vegetables, fats, grains, dairy, beverages, and processed foods were subjected to

monosaccharide analysis and the resulting monosaccharide compositions were used to create a

"Glycopedia." The results from the Glycopedia can be used to determine specific

monosaccharide amounts in foods and is capable of tailoring diets. Correlations were made

within monosaccharides found in food. Clustering analysis was performed to see correlation in

food groups based on quantitative monosaccharide compositions and amounts. Using this LC-

MS method, we determined the monosaccharide composition of foods. The resulting information

provided for a more precise diet tailored to modulate the gut microbiome, which could lead to

better health.

REFERENCES

1.   Le Couteur, D. G.; Solon-Biet, S.; Wahl, D.; Cogger, V. C.; Willcox, B. J.; Willcox, D. C.; Raubenheimer, D.; Simpson, S. J., New Horizons: Dietary protein, ageing and the Okinawan ratio. *Age Ageing* **2016,** *45* (4), 443-7.
2.   Henrick, B. M.; Rodriguez, L.; Lakshmikanth, T.; Pou, C.; Henckel, E.; Arzoomand, A.; Olin, A.; Wang, J.; Mikes, J.; Tan, Z.; Chen, Y.; Ehrlich, A. M.; Bernhardsson, A. K.; Mugabo, C. H.; Ambrosiani, Y.; Gustafsson, A.; Chew, S.; Brown, H. K.; Prambs, J.; Bohlin, K.; Mitchell, R. D.; Underwood, M. A.; Smilowitz, J. T.; German, J. B.; Frese, S. A.; Brodin, P., Bifidobacteria-mediated immune system imprinting early in life. *Cell* **2021,** *184* (15), 3884-3898 e11.
3.   Collaborators, G. B. D. D., Health effects of dietary risks in 195 countries, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* **2019,** *393* (10184), 1958-1972.
4.   Cordain, L.; Eaton, S. B.; Sebastian, A.; Mann, N.; Lindeberg, S.; Watkins, B. A.; O'Keefe, J. H.; Brand-Miller, J., Origins and evolution of the Western diet: health implications for the 21st century. *Am J Clin Nutr* **2005,** *81* (2), 341-54.
5.   Liu, L.; Wang, S.; Liu, J., Fiber consumption and all-cause, cardiovascular, and cancer mortalities: a systematic review and meta-analysis of cohort studies. *Mol Nutr Food Res* **2015,** *59* (1), 139-46.
6.   Sonnenburg, J. L.; Backhed, F., Diet-microbiota interactions as moderators of human metabolism. *Nature* **2016,** *535* (7610), 56-64.
7.   Makki, K.; Deehan, E. C.; Walter, J.; Backhed, F., The Impact of Dietary Fiber on Gut Microbiota in Host Health and Disease. *Cell Host Microbe* **2018,** *23* (6), 705-715.
8.   Amicucci, M. J.; Nandita, E.; Galermo, A. G.; Castillo, J. J.; Chen, S.; Park, D.; Smilowitz, J. T.; German, J. B.; Mills, D. A.; Lebrilla, C. B., A nonenzymatic method for cleaving polysaccharides to yield oligosaccharides for structural analysis. *Nat Commun* **2020,** *11* (1), 3963.
9.   Phillips, K. M.; Haytowitz, D. B.; Pehrsson, P. R., Implications of two different methods for analyzing total dietary fiber in foods for food composition databases. *J Food Compos Anal* **2019,** *84,* 103253.
10.   Barratt, M. J.; Lebrilla, C.; Shapiro, H. Y.; Gordon, J. I., The Gut Microbiota, Food Science, and Human Nutrition: A Timely Marriage. *Cell Host Microbe* **2017,** *22* (2), 134-141.
11.   U.S. Department of Agriculture, U. S. D. o. H. a. H. S., Dietary Guidelines for Americans, 2020-2025. 9th Edition. **2020**.
12.   Baldiviez, L. M.; Keim, N. L.; Laugero, K. D.; Hwang, D. H.; Huang, L.; Woodhouse, L. R.; Burnett, D. J.; Zerofsky, M. S.; Bonnel, E. L.; Allen, L. H.; Newman, J. W.; Stephensen, C. B., Design and implementation of a cross-sectional nutritional phenotyping study in healthy US adults. *BMC Nutr* **2017,** *3*, 79.
13.   Service, U. A. R. What We Eat in America (WWEIA) Food Categories.
14.   Service, U. A. R. Food and Nutrient Database for Dietary Studies (FNDDS) 2017-2018.
15.   Sciences, N. D. o. C. C. a. P. Automated Self-Administered 24-Hour (ASA24) Dietary Assessment Tool.

16.   Xu, G.; Amicucci, M. J.; Cheng, Z.; Galermo, A. G.; Lebrilla, C. B., Revisiting monosaccharide analysis - quantitation of a comprehensive set of monosaccharides using dynamic multiple reaction monitoring. *Analyst* **2017,** *143* (1), 200-207.

17.   Amicucci, M. J.; Galermo, A. G.; Nandita, E.; Vo, T. T. T.; Liu, Y. Y.; Lee, M.; Xu, G. G.; Lebrilla, C. B., A rapid-throughput adaptable method for determining the monosaccharide composition of polysaccharides. *Int J Mass Spectrom* **2019,** *438*, 22-28.

18.   Service, U. A. R. Food and Nutrient Database for Dietary Studies (FNDDS) 2015-2016.

19.   Mohnen, D., Pectin structure and biosynthesis. *Curr Opin Plant Biol* **2008,** *11* (3), 266-77.

20.   Burton, R. A.; Fincher, G. B., Evolution and development of cell walls in cereal grains. *Front Plant Sci* **2014,** *5*, 456.

21.   Delannoy-Bruno, O.; Desai, C.; Raman, A. S.; Chen, R. Y.; Hibberd, M. C.; Cheng, J. Y.; Han, N.; Castillo, J. J.; Couture, G.; Lebrilla, C. B.; Barve, R. A.; Lombard, V.; Henrissat, B.; Leyn, S. A.; Rodionov, D. A.; Osterman, A. L.; Hayashi, D. K.; Meynier, A.; Vinoy, S.; Kirbach, K.; Wilmot, T.; Heath, A. C.; Klein, S.; Barratt, M. J.; Gordon, J. I., Evaluating microbiome-directed fibre snacks in gnotobiotic mice and humans. *Nature* **2021,** *595* (7865), 91-95.

22.   Marcotuli, I.; Hsieh, Y. S.; Lahnstein, J.; Yap, K.; Burton, R. A.; Blanco, A.; Fincher, G. B.; Gadaleta, A., Structural Variation and Content of Arabinoxylans in Endosperm and Bran of Durum Wheat (Triticum turgidum L.). *J Agric Food Chem* **2016,** *64* (14), 2883-92.

23.   Yatsunenko, T.; Rey, F. E.; Manary, M. J.; Trehan, I.; Dominguez-Bello, M. G.; Contreras, M.; Magris, M.; Hidalgo, G.; Baldassano, R. N.; Anokhin, A. P.; Heath, A. C.; Warner, B.; Reeder, J.; Kuczynski, J.; Caporaso, J. G.; Lozupone, C. A.; Lauber, C.; Clemente, J. C.; Knights, D.; Knight, R.; Gordon, J. I., Human gut microbiome viewed across age and geography. *Nature* **2012,** *486* (7402), 222-227.

24.   Henrick, B. M.; Chew, S.; Casaburi, G.; Brown, H. K.; Frese, S. A.; Zhou, Y.; Underwood, M. A.; Smilowitz, J. T., Colonization by B. infantis EVC001 modulates enteric inflammation in exclusively breastfed infants. *Pediatr Res* **2019,** *86* (6), 749-757.

25.   Buffie, C. G.; Pamer, E. G., Microbiota-mediated colonization resistance against intestinal pathogens. *Nat Rev Immunol* **2013,** *13* (11), 790-801.

26.   Arzamasov, A. A.; van Sinderen, D.; Rodionov, D. A., Comparative Genomics Reveals the Regulatory Complexity of Bifidobacterial Arabinose and Arabino-Oligosaccharide Utilization. *Front Microbiol* **2018,** *9,* 776.

27.   Schell, M. A.; Karmirantzou, M.; Snel, B.; Vilanova, D.; Berger, B.; Pessi, G.; Zwahlen, M. C.; Desiere, F.; Bork, P.; Delley, M.; Pridmore, R. D.; Arigoni, F., The genome sequence of Bifidobacterium longum reflects its adaptation to the human gastrointestinal tract. *P Natl Acad Sci USA* **2002,** *99* (22), 14422-14427.

28.   Cantarel, B. L.; Coutinho, P. M.; Rancurel, C.; Bernard, T.; Lombard, V.; Henrissat, B., The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res* **2009,** *37* (Database issue), D233-8.

29.   Akbari, P.; Braber, S.; Alizadeh, A.; Verheijden, K. A.; Schoterman, M. H.; Kraneveld, A. D.; Garssen, J.; Fink-Gremmels, J., Galacto-oligosaccharides Protect the Intestinal Barrier by Maintaining the Tight Junction Network and Modulating the Inflammatory Responses after a Challenge with the Mycotoxin Deoxynivalenol in Human Caco-2 Cell Monolayers and B6C3F1 Mice. *J Nutr* **2015,** *145* (7), 1604-13.

30.   Natividad, J. M.; Rytz, A.; Keddani, S.; Bergonzelli, G.; Garcia-Rodenas, C. L., Blends of Human Milk Oligosaccharides Confer Intestinal Epithelial Barrier Protection in Vitro. *Nutrients* **2020,** *12* (10).

31.   Li, F.; Du, P.; Yang, W.; Huang, D.; Nie, S.; Xie, M., Polysaccharide from the seeds of Plantago asiatica L. alleviates nonylphenol induced intestinal barrier injury by regulating tight junctions in human Caco-2 cell line. *Int J Biol Macromol* **2020,** *164*, 2134-2140.

32.   Li, Y.; Tian, X.; Li, S.; Chang, L.; Sun, P.; Lu, Y.; Yu, X.; Chen, S.; Wu, Z.; Xu, Z.; Kang, W., Total polysaccharides of adlay bran (Coix lachryma-jobi L.) improve TNF-alpha induced epithelial barrier dysfunction in Caco-2 cells via inhibition of the inflammatory response. *Food Funct* **2019,** *10* (5), 2906-2913.

33.   Lu, Y.; Li, L.; Zhang, J. W.; Zhong, X. Q.; Wei, J. A.; Han, L., Total polysaccharides of the Sijunzi decoction attenuate tumor necrosis factor-alpha-induced damage to the barrier function of a Caco-2 cell monolayer via the nuclear factor-kappaB-myosin light chain kinase-myosin light chain pathway. *World J Gastroenterol* **2018,** *24* (26), 2867-2877.

34.   Barnett, A. M.; Roy, N. C.; McNabb, W. C.; Cookson, A. L., Effect of a Semi-Purified Oligosaccharide-Enriched Fraction from Caprine Milk on Barrier Integrity and Mucin Production of Co-Culture Models of the Small and Large Intestinal Epithelium. *Nutrients* **2016,** *8* (5).

35.   Anderson, J. W.; Baird, P.; Davis, R. H., Jr.; Ferreri, S.; Knudtson, M.; Koraym, A.; Waters, V.; Williams, C. L., Health benefits of dietary fiber. *Nutr Rev* **2009,** *67* (4), 188-205.

36.   Gonzalez, P. S.; O'Prey, J.; Cardaci, S.; Barthet, V. J. A.; Sakamaki, J. I.; Beaumatin, F.; Roseweir, A.; Gay, D. M.; Mackay, G.; Malviya, G.; Kania, E.; Ritchie, S.; Baudot, A. D.; Zunino, B.; Mrowinska, A.; Nixon, C.; Ennis, D.; Hoyle, A.; Millan, D.; McNeish, I. A.; Sansom, O. J.; Edwards, J.; Ryan, K. M., Mannose impairs tumour growth and enhances chemotherapy. *Nature* **2018,** *563* (7733), 719-723.

37.   Rajoka, M. S. R.; Shi, J. L.; Mehwish, H. M.; Zhu, J.; Li, Q.; Shao, D. Y.; Huang, Q. S.; Yang, H., Interaction between diet composition and gut microbiota and its impact on gastrointestinal tract health. *Food Sci Hum Well* **2017,** *6* (3), 121-130.

38.   B.C. Tungland, D. M., Nondigestible Oligo- and Polysaccharides (Dietary Fiber): Their Physiology and Role in Human Health and Food. *Comprehensive Reviews in Food Science and Food Safety* **2006,** *1* (3), 90-109.

Chapter 3

A Multidimensional Mass Spectrometry-based Workflow for *de Novo* Structural Elucidation of Oligosaccharides from Polysaccharides

ABSTRACT

Carbohydrates play essential roles in a variety of biological processes that are dictated by their structures. However, characterization of carbohydrate structures remains extremely difficult and generally unsolved. In this work, a *de novo* mass spectrometry-based workflow was developed to isolate and structurally elucidate oligosaccharides to provide sequence, monosaccharide compositions, and glycosidic linkage positions. The approach employs liquid chromatography−tandem mass spectrometry (LC−MS/MS)-based methods in a 3-dimensional (3D) concept: one high performance liquid chromatography-quadrupole time-of-flight mass spectrometry (HPLC-QTOF MS) analysis for oligosaccharide sequencing and two ultra high performance liquid chromatography-triple quadrupole mass spectrometry (UHPLC-QqQ MS) analyses on fractionated oligosaccharides to determine their monosaccharide and linkage compositions. The workflow was validated by applying the procedure to maltooligosaccharide standards. The approach was then used to determine the structures of oligosaccharides derived from polysaccharide standards and whole food products. The integrated LC-MS workflow will reveal the in-depth structures of oligosaccharides.

INTRODUCTION

Carbohydrates have a wide range of important functions that are dictated by their specific structures. For example, human milk oligosaccharides (HMOs) are not digested by the infant's innate metabolism. However, they are catabolized by the infant gut bacteria thereby providing nutritional value in the form of short chain fatty acids to the infant.[1,2] The specific structures of HMOs correspond uniquely to bacterial enzymes that ferment them.[3–10] In human cells, oligosaccharides are also found on cell surfaces where they play important roles in cell to cell recognition and cell binding processes.[11–14] Under some conditions, gut bacteria are known to consume host cell glycans in a similarly structure-specific manner.[15] In food, carbohydrates are often the largest component of the adult diet and are similarly responsible for shaping the composition of the gut microbiota.[16,17]

Dietary oligosaccharides such as fructooligosaccharides (FOS) and galactooligosaccharides (GOS) are termed prebiotics because they are similarly not digested by the host, but are consumed by gut bacteria.[18–20] These oligosaccharides are similarly selectively fermented by microbes in a structure specific manner in the large intestine.[21,22] While FOS and GOS represent major industrial products, they are relatively homogenous and are composed primarily of only a single monomer (fructose or galactose). Very many plant-based oligosaccharides may already be accessible to gut bacteria, however we currently lack the methods to produce and characterize them. Methods for the *de novo* structural elucidation of oligosaccharides and polysaccharides are critically needed to develop new classes of prebiotics with potentially higher biological efficacy.

The development of methods for the complete analysis of oligosaccharides presents major challenges. A variety of techniques are typically employed to determine different aspects of their structures. Size exclusion chromatography (SEC) is used to separate oligosaccharides from complex matrices and gives insight into the average degree of polymerization, but does not elucidate the identity of the monosaccharides nor glycosidic linkages.[23] Other methods including capillary electrophoresis (CE), high performance liquid chromatography (HPLC), and high performance anion-exchange chromatography (HPAEC) are used to separate oligosaccharides, while ultraviolet-visible (UV-Vis) spectrophotometers, refractive index detectors (RID) and pulsed amperometric detectors (PAD) are used for detection.[24–27] While these methods are sensitive, they lack throughput, structural information, and integration in a way that allows a deep characterization of structure. Gas chromatography (GC) is typically less used for the analysis of oligosaccharides, but has long been used as the standard method for monosaccharide and linkage analysis.[28,29] GC methods, however require long separation times, have low sensitivities and throughput, require extra derivatization steps in the sample preparation, focus on a limited number of monosaccharides and linkages, and yield little to no absolute quantitative information.[30–32] Mass spectrometry method such as collision-induced dissociation (CID) have been employed with sodiated oligosaccharides to yield specific structural information.[33–35] Newer MS techniques, such as ion mobility mass spectrometry (IM-MS) while still being developed could potentially provide anomer and sequence information. However, these and similar methods do not yet provide *de novo* monosaccharide nor linkage information.[36–38]

The impact that carbohydrates have on the microbiota has renewed interests in developing analytical methods for poly- and oligosaccharide analysis.[39–43] The understanding of the role of HMOs, for example, has increased when we developed methods to obtain complete structures of

large pools of compounds through LC-MS profiling. Tandem mass spectrometry (LC-MS/MS) further revealed structural information such as the degree of polymerization (DP) with composition and some sequence information when mass heterogeneity of the monomers are available.[44,45] However, challenges in oligosaccharide analysis using LC-MS/MS remain. Many monosaccharides have identical masses so that CID fragmentation is insufficient for differentiating the isomers or branching positions. A hexose monosaccharide with a single mass can have several isomers which include fructose, glucose, galactose, mannose, or allose, and cannot be resolved using standard LC-MS/MS workflow. Enzymatic treatments coupled with LC-MS/MS have been employed to identify the monosaccharide isomers and yield complete oligosaccharide structures.[46] Enzymes used in this way were effective, however they require prior knowledge of the monosaccharide compositions, possible linkages, and further require commercial availability.

In this report, we describe a workflow for *de novo* structural elucidation of oligosaccharides derived from plant polysaccharides. The concept involves three types of LC-MS/MS analyses representing a 3-Dimensional mass spectrometry (3D-MS) approach to reveal each oligosaccharide's sequence, composition, and linkages. HPLC separation of the oligosaccharides is performed followed by a two-way split to an in-line fractionation and a high-resolution quadrupole time of flight mass spectrometry instrument. A 96-well plate fraction collector is used to isolate the oligosaccharides. Structural information regarding the DP is obtained through tandem MS (MS/MS). The collected fractions are subjected to a series of rapid-throughput methods to provide monosaccharide and linkage profiles.[41,47,48] For each oligosaccharide, the method yields absolute quantitation of the monosaccharides and the relative abundance of each glycosidic linkage. Using this combinatorial approach that comprises sequence information from the QTOF

MS/MS and the monosaccharide and linkage compositions from separate LC-MS/MS analyses, the in-depth oligosaccharide structures of the mixtures are elucidated.

EXPERIMENTAL SECTION

Samples and materials. Sodium acetate, hydrogen peroxide (H_2O_2) (30% v/v), iron(III) sulfate pentahydrate ($Fe_2(SO_4)_3$), trifluoroacetic acid (TFA) (HPLC grade), chloroform (HPLC grade), ammonium acetate, sodium hydroxide pellets (semi-conductor grade 99.99% trace metals basis), ammonium hydroxide solution (NH_4OH) (28-30%), 3-methyl-1-phenyl-2-pyrazoline-5-one (PMP), dichloromethane (DCM), methanol (HPLC grade), anhydrous dimethyl sulfoxide (DMSO), glacial acetic acid, D-galactose, D-mannose, D-glucose, D-allose, D-fructose, D-xylose, L-arabinose, D-xylose, L-fucose, L-rhamnose, N-acetyl-D-glucosamine (GlcNAc), N-acetyl-D-galactosamine (GalNAc), D-glucuronic acid (GlcA), and D-galacturonic acid (GalA) were purchased from Sigma-Aldrich (St. Louis, MO). Maltotriose, maltotetraose, maltopentaose, maltohexaose were obtained from Carbosynth (Compton, UK). Galactan was purchased from Megazyme (Bray, Ireland). C18 and porous graphitized carbon (PGC) solid phase extraction (SPE) plates were purchased from Glygen (Columbia, MD). Formic Acid (FA) (99.5% optima LC/MS grade) was purchased from Fisher Scientific (Hampton, NH). Acetonitrile (ACN) (HPLC grade) was purchased from Honeywell (Muskegon, MI). All purchased chemicals were used without further purification. Butternut squash was purchased from Pedrick Produce (Dixon, CA). Nanopure water was used for all experiments.

Preparation of the maltooligosaccharides. A stock solution comprising of maltotriose, maltotetraose, maltopentaose, and maltohexaose were prepared in water at a concentration of 2 mg/mL. A 10 uL aliquot was subjected to the 3D MS workflow.

Preparation of polysaccharide derived oligosaccharides. An oxidative digestion step, Fenton's initiation towards defined oligosaccharide groups (FITDOG) was used to generate oligosaccharides from polysaccharides.[49] Briefly, a reaction solution was prepared containing 95% (v/v) 40 mM sodium acetate buffer adjusted to pH 5 with glacial acetic acid, 5% (v/v) hydrogen peroxide (30% v/v), and 65 nM iron(III) sulfate. The reaction mixture was added to 1 mg of dry galactan, and 1 mg lyophilized butternut squash to make a final solution of 1 mg/mL of both. The reaction was incubated at 100°C for one hour. The reaction was quenched by adding half of the reaction volume of cold 2 M NaOH. Glacial acetic acid was added for neutralization. A total of thirty reactions were performed on a 96-well plate.

Oligosaccharides generated from galactan were purified by 96-well plate PGC cartridges. The cartridges were washed with 80% acetonitrile and 0.1% (v/v) TFA in water. The oligosaccharides were loaded and washed with five column volumes of water. The oligosaccharides were eluted with 40% acetonitrile with 0.05% (v/v) TFA in water. Samples were pooled, dried by vacuum centrifugation, and reconstituted in water prior to the *de novo* oligosaccharide analysis.

An additional C18 clean up step was employed to oligosaccharides generated from butternut squash before the PGC cartridge enrichment. For the C18 cleanup, cartridges were washed with pure acetonitrile followed a water wash five times the columns volumes. Analytes were loaded and eluted with two column volumes of pure water. The effluent containing the oligosaccharides were then enriched using the PGC protocol described above.

Mass spectrometry analysis.

Fractionation of oligosaccharides. The oligosaccharide separation was performed on an Agilent 1260 Infinity II series HPLC. Inline fractionation was performed on a two-way flow splitter coupled to an Agilent 6530 QTOF mass spectrometer and a Teledyne Isco Foxy 200 96-well plate fraction collector. A 150 mm x 4.6 mm Hypercarb column from Thermo Scientific with a 5 µm particle size was used to chromatographically separate the oligosaccharides. A binary gradient consisting of solvent A: (3 % (v/v) acetonitrile/water + 0.1 % (v/v) formic acid) and solvent B: (90 % (v/v) acetonitrile/water + 0.1 % (v/v) formic acid) was used. A 45-minute gradient with a flow rate of 1 mL/min was used for the separation of maltooligosaccharides 0.00-30.00 mins, 8.00-10.00 % B; 30.00-30.01 mins; 10.00-99.00 % B; 30.01-33.00 mins, 99.00 % B; 33.00-33.01 mins, 99.00-8.00 % B, 33.01-45.00 mins, 8.00 % B. A 120-minute gradient with a flow rate of 1 mL/min was used for chromatographic separation of oligosaccharides generated from galactan and butternut squash: 0.00-90.00 mins, 5.00-12.00 % B; 90.00-99.01 mins, 12.00-99.00 % B; 90.01-110.00 mins, 99.00-99.00 % B; 110.00-110.01 mins, 99.00-5.00 % B; 110.01-120.00 mins, 5.00 % B.

A two-way flow splitter was used to divert 90 % of the effluent to the fraction collector and the remaining 10 % of the effluent to the QTOF mass spectrometer. Fractions were collected on 96-well plates at a rate of 0.5 min/fraction. The fractions were then dried to completion under vacuum centrifugation and reconstituted in 100 µL of nano-pure water. A 10 µL aliquot was transferred to a separate 96-well plate for monosaccharide composition analysis, while the remaining 90 µL of solution was used for glycosidic linkage analysis. Both 96-well plates were then dried to completeness by vacuum centrifugation prior to monosaccharide and linkage analysis.

Sample was introduced into the QTOF instrument using an electrospray ionization (ESI) source operated in positive mode. The instrument was calibrated with internal calibrant ions ranging from m/z 118.086-2721.895. Drying gas was set to 150 °C and with a flow rate of 11 L/min. The fragment, skimmer, and Octapole 1 RF voltages were set to 75, 60, and 750 V, respectively. The collision energy was based upon the compound mass and expressed by the linear function (Collision Energy $=1.3*(m/z)$-3.5). The QTOF data was acquired using Agilent MassHunter Workstation Data Acquisition version B.08.00. The data was then further analyzed by Agilent MassHunter Qualitative Analysis software version B.07.00.

Monosaccharide analysis of fractions. Monosaccharide compositional analysis performed on separated oligosaccharides used methods from Amicucci et al.[47] and Xu et al.[48]. The separated oligosaccharides underwent a hard acid hydrolysis with 4 M TFA for 2 hours at 100°C. The released monosaccharides were dried to completion by vacuum centrifugation. A calibration curve consisting of monosaccharides standards containing D-glucose, D-galactose, D-mannose, D-fructose, D-allose, L-fucose, L-rhamnose, L-arabinose, D-xylose, D-ribose, GlcA, GalA, GlcNAc, and GalNAc were prepared in water at concentrations ranging from 0.001 µg/µL to 100 µg/µL. Samples and monosaccharide standards were derivatized with 0.2 M PMP dissolved in methanol and 28% NH_4OH at 70°C for 30 minutes. Derivatized products were dried to completion under vacuum centrifugation and reconstituted in nano-pure water. Excess PMP was removed with chloroform extraction. A 1 µL aliquot from the aqueous layer was analyzed by an Agilent 1290 Infinity II UHPLC coupled to an Agilent 6495A QqQ MS employing dynamic multiple reaction monitoring (dMRM) mode. An external standard curve was used for absolute quantitation of each monosaccharide in the fraction. The data analysis was performed using

Agilent MassHunter Qualitative Analysis software version B.07.00. and Agilent MassHunter Quantitative Analysis software version B.05.02.

Linkage analysis of fractions. Linkage analysis performed on separated oligosaccharides used the method from Galermo et al.[41] In short, permethylation was performed on separated oligosaccharides by reacting the compounds with 5 µL of 1.26 g/mL saturated aqueous NaOH and 150 µL of DMSO. A 40 uL aliquot of iodomethane was added to the mixture and reacted under argon on a shaker at room temperature for 50 minutes. Residual NaOH and DMSO were removed by extraction with DCM and water. The DCM layer was dried to completion by vacuum centrifugation. Permethylated oligosaccharides were hydrolyzed and derivatized in the same manner as the monosaccharide analysis. After PMP derivatization step and the drying step, the permethylated glycosides were reconstituted in 100 µL of 70% (v/v) methanol/water. A 1 µL aliquot was analyzed on an Agilent 1290 Infinity II UHPLC coupled to an Agilent 6495A QqQ MS instrument in multiple reaction monitoring (MRM) mode. The data analysis was performed using Agilent MassHunter Qualitative Analysis software version B.07.00. and Agilent MassHunter Quantitative Analysis software version B.05.02. A pool of oligosaccharide standards was used to assign linkages.

RESULTS AND DISCUSSION

Description of the overall workflow. A 3D mass spectrometry workflow was developed to structurally elucidate oligosaccharides from polysaccharides using a combination of high resolution and targeted mass spectrometry methods. This approach incorporates a non-enzymatic digestion step to generate polysaccharide-based oligosaccharides followed by three mass spectrometry analyses to obtain the in-depth structural characterization of oligosaccharides. Oligosaccharides were subjected to the workflow shown in **Figure 3.1**. The oligosaccharides are

first subjected to HPLC separation using a Hypercarb PGC column. During the chromatographic

run, the effluent was split two ways, one toward a QTOF MS and the other toward a 96-well

plate fraction collector. The isolated oligosaccharides were characterized in the QTOF by their

respective MS and MS/MS spectra. To further determine the monosaccharides and linkages of

the separated oligosaccharides, collected fractions were subjected to monosaccharide analysis

and linkage analysis. The workflow is the first to use a combination of three LC-MS platforms to

elucidate the in-depth structures of oligosaccharides. Furthermore, it provides several

advantages, namely higher throughput in the monosaccharide and linkage determination

employing a 96-well plate format.
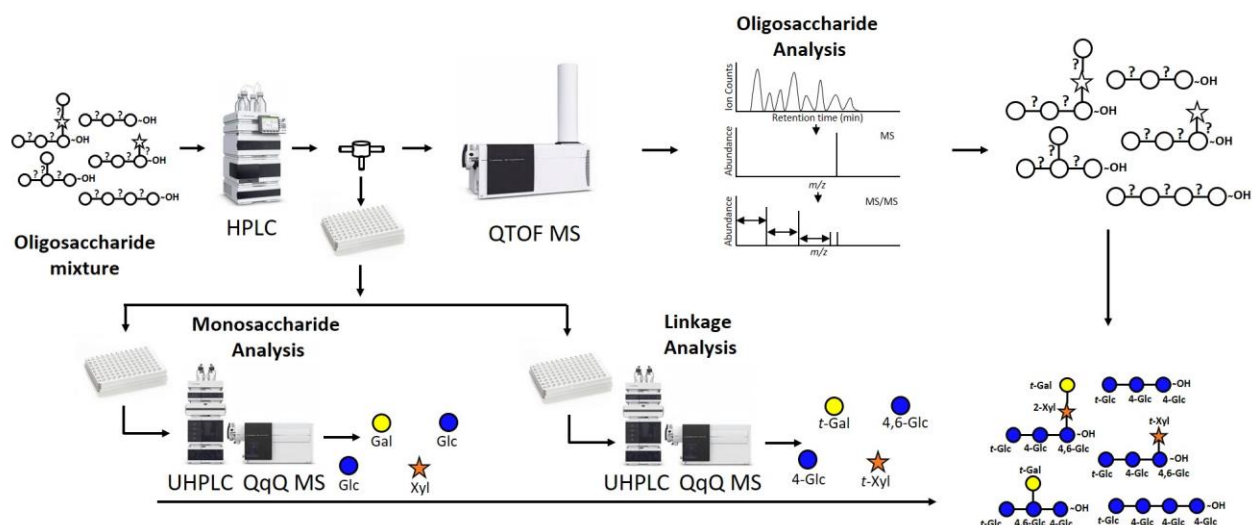
Figure 3.1. Workflow for the *de novo* structural elucidation of oligosaccharides using the 3D

LC-MS/MS. Samples are injected in the LC-QTOF MS. The eluant is split using a flow splitter

between the LC and the QTOF MS to a 96-well plate collector. The collected oligosaccharide

fractions are subjected for UHPLC-QqQ MS analyses to determine monosaccharide and linkage

compositions.

Workflow validation using maltooligosaccharides. The workflow was demonstrated using an oligosaccharide mixture comprised of maltotriose, maltotetraose, maltopentaose, and maltohexaose. The maltooligosaccharides were chosen to model plant-based oligosaccharides derived from polysaccharides. A stock solution of the standards were pooled and subjected to the workflow. To isolate the oligosaccharides, chromatographic conditions were optimized to yield a 45-minute run while still achieving base peak separation for the oligosaccharides. Complete chromatographic separation of the oligosaccharides was preferred, yet challenging to achieve, as many oligosaccharides had similar structures that tend to coelute in the more complicated mixtures described below.

The first dimension of MS results in the workflow was oligosaccharide analysis obtained through LC-QTOF MS. Two scan modes were employed on the oligosaccharides: a precursor ion scan mode to identify possible oligosaccharides based on their corresponding *m/z* value and a fragmentation mode, where collision-induced dissociation (CID) was employed to selected oligosaccharide ions. Fragmentation of oligosaccharide ions yielded structural information. In general, native oligosaccharide ions fragmented under CID inducing cleavages at the glycosidic bonds. Depending on the monosaccharide type (a hexose versus a pentose or a deoxyhexose), oligosaccharides can produce unique fragmentation spectra. For example, hexose, pentose, and a deoxyhexose yield losses of 162, 132, and 146 Da, respectively. The neutral mass differences and the resulting fragment ions are used to determine the monosaccharide class, some spatial arrangements, and the DP. We further note that rearrangements can occur with protonated species during CID, and these rearrangements can complicate structural analysis.

For the standards, the LC-QTOF MS chromatogram revealed a total of four oligosaccharides in the mixture shown in **Figure 3.2A**. The MS and MS/MS analysis revealed

the *m/z* 505.178 as a trisaccharide, *m/z* 667.231 tetrasaccharide, *m/z* 829.283 pentasaccharide, and *m/z* 991.331 hexasaccharide. The compounds were split by the alpha and beta anomer content except for the trisaccharide, which did not interact as strongly on the stationary phase. The DP of the oligosaccharides and the monosaccharide compositions (hexoses) were readily identified in the standards.
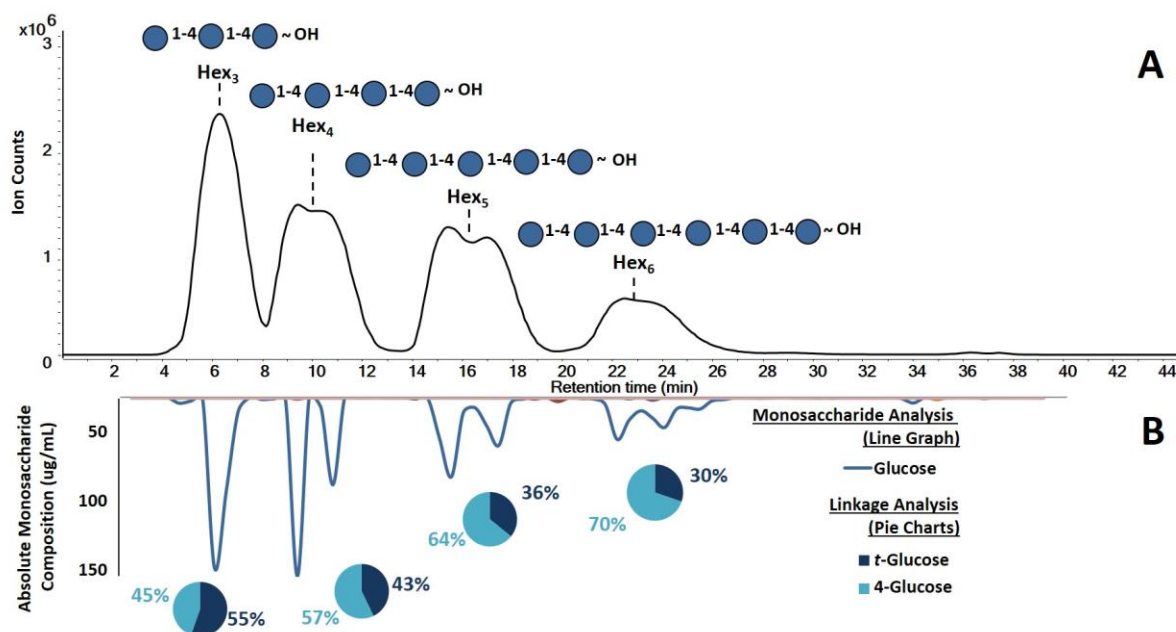
Figure 3.2. Elucidated maltooligosaccharide structures obtained from 3D LC-MS/MS workflow. (A) The 1st dimension corresponds to base peak chromatogram (BPC) to yield the oligosaccharide profiles. (B) The 2nd dimension revealed the absolute monosaccharides abundances in the fractions (inverted solid line graph). The 3rd dimension produced the glycosidic linkages within the oligosaccharide (Inset, pie charts).

To illustrate the details of the oligosaccharide analysis, maltohexaose was used. In the positive mode, the precursor ion corresponded to the mass of the protonated species $[M+H]^+$

(991.331 Da, **Figure 3.3A**). Maltohexaose was subjected to CID fragmentation at a collision energy of 9.4 eV and yielded abundant fragment ions revealing the connectivity and confirming the DP (**Figure 3.3B**). The fragment ions m/z 163.062 (B_1), m/z 325.113 (B_2), m/z 487.166 (B_3), m/z 649.212 (B_4), and m/z 811.272 (B_5) corresponded to one to five hexose residue(s), respectively. The observed fragment ions were consistent with the known structure of maltohexaose.
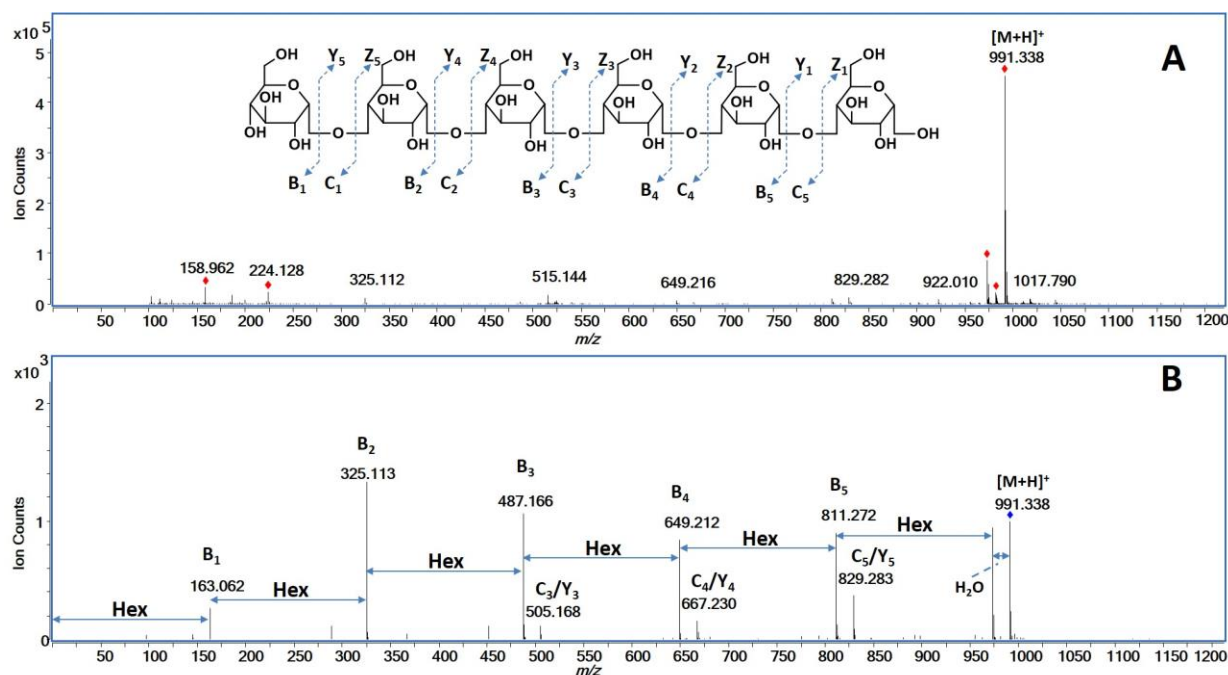
Figure 3.3. (A) MS and (B) MS/MS spectra of protonated maltohexaose species [M+H]$^+$ with assigned fragmentation obtained from the LC-QTOF MS.

The next dimension of MS in the 3D workflow yielded the monosaccharide compositional analysis. This analysis was performed on the collected fractions that were split from the LC-MS and collected in 96-well plates. A total of 76 fractions were collected and subjected to acid hydrolysis to produce monosaccharides. The released monosaccharides were then reacted in a PMP derivatization and subsequently analyzed by UHPLC-QqQ MS. The LC-

MS monosaccharide analysis was performed in 10 minutes. For the standard pool, the monosaccharide results were plotted as an inverted chromatogram shown in **Figure 3.2B**. The chromatogram was constructed by plotting the absolute monosaccharide concentrations in each collected fraction with line smoothing performed to produce the figure. The compositional analysis included 14 monosaccharides; however only glucose was found as expected.

The absolute quantitation of monosaccharides was used to validate the fractionation of the oligosaccharides. Recovery analysis was performed on the LC of the maltooligosaccharide pool. LC fractions were analyzed for glucose content using the method for monosaccharide analysis. The total amounts of glucose in the fractions were compared to measured values. Fractional recoveries were obtained using the absolute amounts measured in each fraction (**Table 3.1**). The glucose recoveries for maltotriose, maltotetraose, maltopentaose, and maltohexaose were determined to be 69.0%, 85.8 %, 89.8%, and 82.5%, respectively. There were high recoveries of the oligosaccharides with variations possibly due to some degradation in the strong acid hydrolysis treatment.

Table 3.1. Percent recoveries of pre and post fractionation of maltooligosaccharide pool based on the total glucose quantitation.

Maltooligosaccharide	Degree of Polymerization (DP)	Collection window (min)	Prefractionated glucose amount (ug)	Fractionated glucose amount (ug)	Glucose recovery (%)
maltotriose	3	4-8	21.9	15.1	69.0
maltotetraose	4	8 -12	16.6	14.2	85.8
maltopentaose	5	14-18	13.3	12.0	89.8
maltohexaose	6	21-26	11.1	9.2	82.5
Total	N/A	N/A	62.9	50.5	80.2

The third MS analysis in the workflow was used to determine the glycosidic linkages. Linkage analysis was performed by permethylating the oligosaccharides, subjecting to acid hydrolysis, and labeling with PMP. The partially methylated compounds were subjected to UHPLC-QqQ MS analysis. The separation for the UHPLC-QqQ MS analysis was performed in 15 minutes for each fraction. This analysis revealed only two types of species: a terminal glucose (*t*-glucose) and 4-linked glucose (4-glucose) for all fractionated oligosaccharides (**Figure 3.2B**). For the oligosaccharides, maltotriose, maltotetraose, maltopentaose, and maltohexaose were observed to contain *t*-glucose and 4-glucose in ratios 1.24, 0.75, 0.56, 0.43. The expected values were 0.50, 0.33, 0.25, 0.20, respectively. The relative abundances for observed and expected values were plotted as a function of DP (**Figure 3.4**) and showed an over-representation of *t*-glucose compared to 4-glucose (**Table 3.2**). The disparity is likely due to specific issues such as differences in the ionization between the various partially methylated species. In general, the MS method may overrepresent terminal linkage species over internal linkage residues. Nonetheless, the analysis generally provided reliable relative abundances of each residue with some quantitative information.
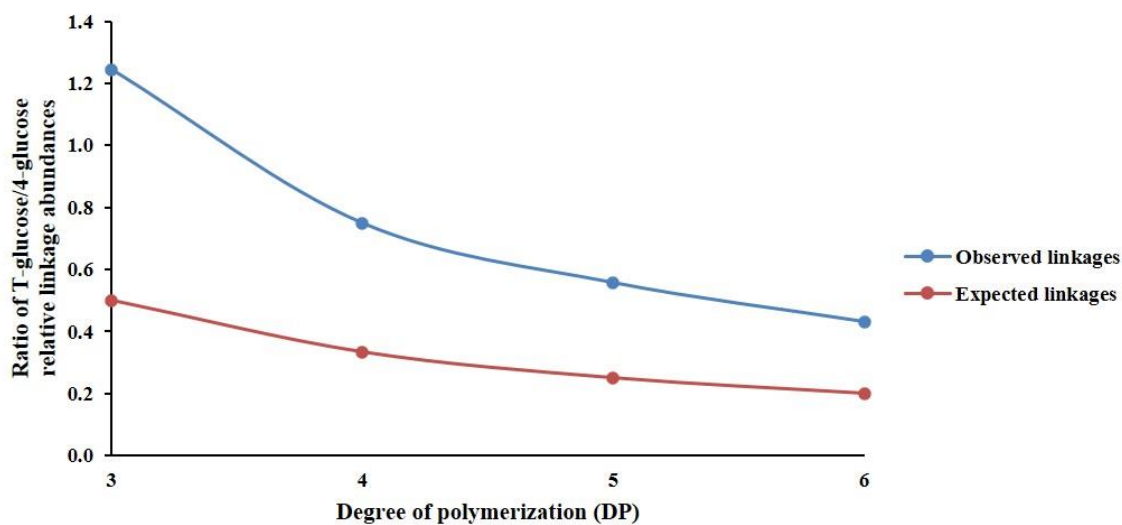
Figure 3.4. Plot of the ratios of T-glucose/4-glucose relative linkage abundances versus DP of observed and expected maltooligosaccharide linkages.

Table 3.2. Degree of polymerization (DP) versus relative linkage composition of observed and expected linkages for maltooligosaccharides.

Maltooligosaccharide	DP	Observed linkages			Expected linkages		
		t-glucose (%)	4-glucose (%)	Ratio (*t*-glucose /4-glucose)	*t*-glucose (%)	4-glucose (%)	Ratio (*t*-glucose /4-glucose)
maltotriose	3	55.4	44.6	1.24	33.3	66.7	0.50
maltotetraose	4	42.8	57.2	0.75	25.0	75.0	0.33
maltopentaose	5	35.8	64.2	0.56	20.0	80.0	0.25
maltohexaose	6	30.2	69.8	0.43	16.7	83.3	0.20

Based on the three LC-MS/MS analyses, the structures of each maltooligosaccharides were deduced. The first QTOF MS with MS/MS, yielded the DPs as 3, 4, 5, and 6. The quantitative UHPLC-QqQ MS analysis revealed that the maltooligosaccharides were solely composed of glucose, while the linkage UHPLC-QqQ MS analysis yielded 4-glucose and *t*-glucose were the only two linkage species observed. The maltooligosaccharide structures were determined to be composed of (1→4) glucose repeating units as shown in the annotated chromatogram in **Figure 3.2A**. While the α/β stereochemical information is currently not obtained with this workflow, future strategies are described below.

Validation of structural analysis of oligosaccharides from standard polysaccharide sources.
A homopolysaccharide galactan was chemically digested using the Fenton's initiation toward defined oligosaccharide groups (FITDOG) process described in previous publication.[49] In this process, polysaccharides were subjected to oxidative cleavage with Fe^{3+}/H_2O_2 conditions. The

process yields distinct oligosaccharides corresponding to each polysaccharide. The resulting

oligosaccharides were subjected to the workflow. The generated oligosaccharides were

chromatographically separated using a 120-minute run and collected into 192 fractions. The

oligosaccharides were then analyzed using Q-TOF MS to determine the DP, hydrolyzed, and

then derivatized to determine the monosaccharide compositions with UHPLC-QqQ MS analysis,

and permethylated, hydrolyzed, and derivatized to determine the glycosidic linkages with

UHPLC-QqQ MS analysis.

The LC-QTOF MS chromatogram of the digested galactan yielded at least seven

dominant oligosaccharides (**Figure 3.5A**). The MS and MS/MS results revealed a Hex_3 (6.5 min,

7.5 min), Hex_4 (12.5 min, 15.0 min), Hex_5 (19.5 min, 22 min), Hex_6 (24.5 min, 27.5 min), Hex_7

(31.5 min, 34.5 min), Hex_8 (45.0 min, 47.0 min), Hex_9 (65.0 min), and Hex_{10} (83.0 min).

Partially modified oligosaccharides were observed and annotated (asterisk) in the inset of **Figure**

3.5A. These non-intact compounds corresponded primarily to higher oxidized products.

Subsequent optimization can eliminate these side products.[49] Separation of anomers (alpha and

beta at the reducing end) resulted in split peaks with the exception of Hex_9 and Hex_{10}. The CID

spectra confirmed the presence of anomers and were further consistent with the proposed

monosaccharide compositions (hexoses).
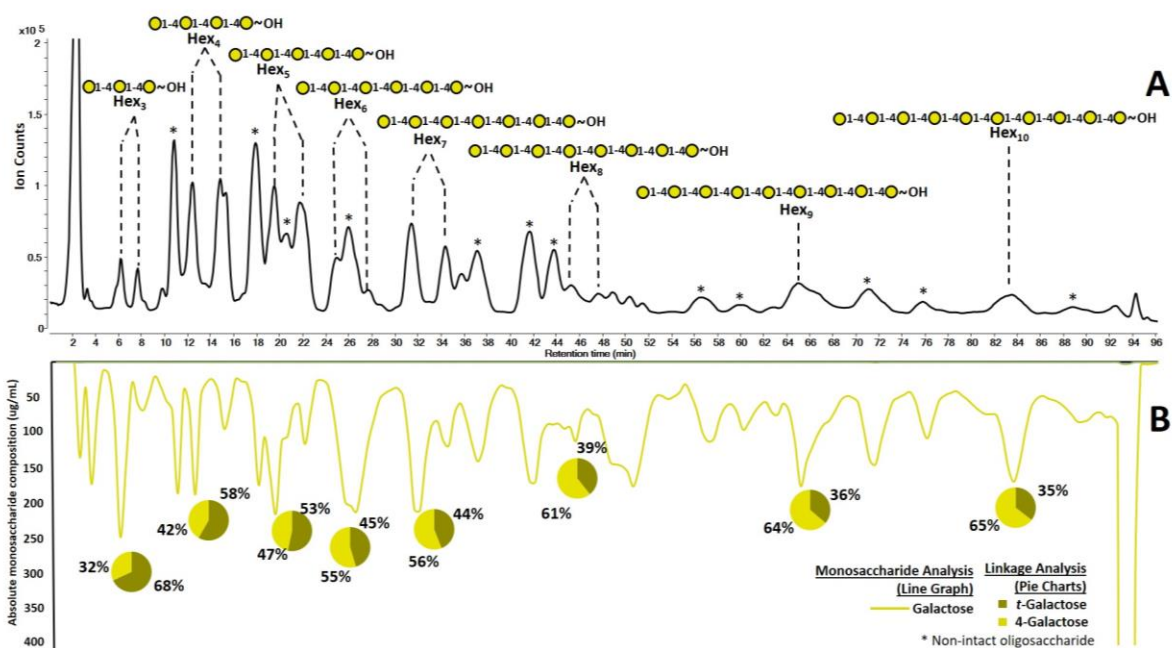
Figure 3.5. Characterization of oligosaccharide structures derived from galactan using the 3D LC-MS. (A) Base peak chromatogram (BPC) from the QTOF MS with the proposed oligosaccharide structures. (B) Mirrored chromatogram represents the monosaccharide composition in fractions and relative linkage compositions are shown in pie charts both obtained through UHPLC-QqQ MS analysis.

The monosaccharide analysis confirmed the composition to be exclusively galactose (**Figure 3.5B**). The linkage analysis similarly yielded only two species, namely 4-galactose and *t*-galactose. The ratios of *t*-galactose and 4-galactose were consistent with the expected ratios, although the terminal linkage species were overrepresented as observed for the maltooligosaccharide (**Table 3.3**). The ratios of *t*-galactose and 4-galactose as function of DP were plotted and were found to follow the trend as with the expected values (**Figure 3.6**). The measured ratios were generally higher than expected, and were significantly higher for the low

DP oligosaccharides suggesting an over representation of *t*-galactose. As mentioned with the maltooligosaccharide example, the terminal species are more abundant than internal linkage species with the effect decreasing with higher DP. A response factor normalized to a polynomial curve may be deduced and used for future measurements if quantitation of the linkages were desired. Nonetheless, the information provided the oligosaccharide structures using the 3D MS analyses. Based on the information, the oligosaccharide structures were determined to be a linear (1→4) linked galactose oligomers as expected for galactan.
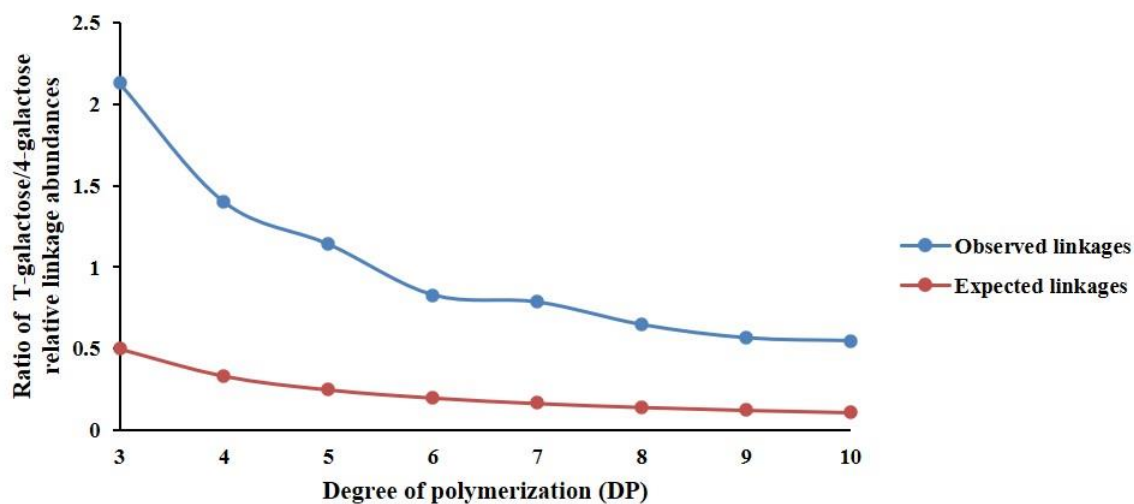
Figure 3.6. Plot of the ratios of T-galactose/4-galactose relative linkage abundances versus DP of observed and expected linkages from galactan oligosaccharides.

Table 3.3. Degree of polymerization (DP) and relative linkage composition of observed and expected linkages for fractionated galactose derived oligosaccharides.

DP	Observed Linkage			Expected Linkage		
	t-galactose (%)	4- galactose (%)	Ratio (*t*- galactose /4- galactose)	*t*- galactose (%)	4- galactose (%)	Ratio (*t*- galactose /4- galactose
3	68.0	32.0	2.13	33.3	66.7	0.50
4	58.3	41.7	1.40	25.0	75.0	0.33
5	53.3	46.7	1.14	20.0	80.0	0.25
6	45.4	54.6	0.83	16.7	83.3	0.20
7	44.0	56.0	0.79	14.3	85.7	0.17
8	39.3	60.7	0.65	12.5	87.5	0.14
9	36.2	63.8	0.57	11.1	88.9	0.13
10	35.5	64.5	0.55	10.0	90.0	0.11

Determination of oligosaccharides produced from food polysaccharides. The 3D MS approach was performed on oligosaccharides produced from a plant product, namely butternut squash. The LC-QTOF chromatogram (**Figure 3.7A**) of the resulting oligosaccharides were separated and collected into a total of 96 fractions. The mirrored chromatogram was constructed from the respective monosaccharide analysis results (**Figure 3.7B**). The resolution of the LC-QTOF was limited by the flow splitter between the LC and the QTOF. Chromatograms without the flow splitter were of significantly higher resolution and were used to confirm the identities of the broad peaks. The major components were composed of DPs varying from disaccharides to pentasaccharides (**Table 3.4**). As expected, a complete plant product such as butternut squash is composed of several polysaccharides. The resulting oligosaccharides yielded a small number of major products namely those composed primarily of hexose units. The monosaccharide analysis of the fractionated effluent yielded the general monosaccharide compositions of the peaks (**Figure 3.7B**). Based on the combined analyses, several compounds were readily identified as those composed of starch. These oligosaccharides consisted of three, four, and five monomer

units (*m/z* 505.180 at 11 minutes, *m/z* 667.235 at 27 minutes, and *m/z* 829.285 at 41 minutes).

Split chromatographic peaks were observed and corresponded to the alpha and beta anomers.

These compounds were comprised almost exclusively of glucose (81.7 %, 90.4 %, 78.1 %,

respectively). The co-elution of other compounds decreased the relative glucose abundances. The

linkage analysis of these compounds yielded 4-glucose (43 %, 59 %, 55 %, respectively) and *t*-

glucose (48 %, 38 %, 39 %) (Pie chart inserts, **Figure 3.7B**). By combining the results from the

three analyses, we conclude that the oligosaccharides were maltotriose, maltotetraose, and

maltopentaose, respectively. Both the retention times and tandem MS matched the

maltooligosaccharide standards, although different LC gradients were used in the separate

analyses.
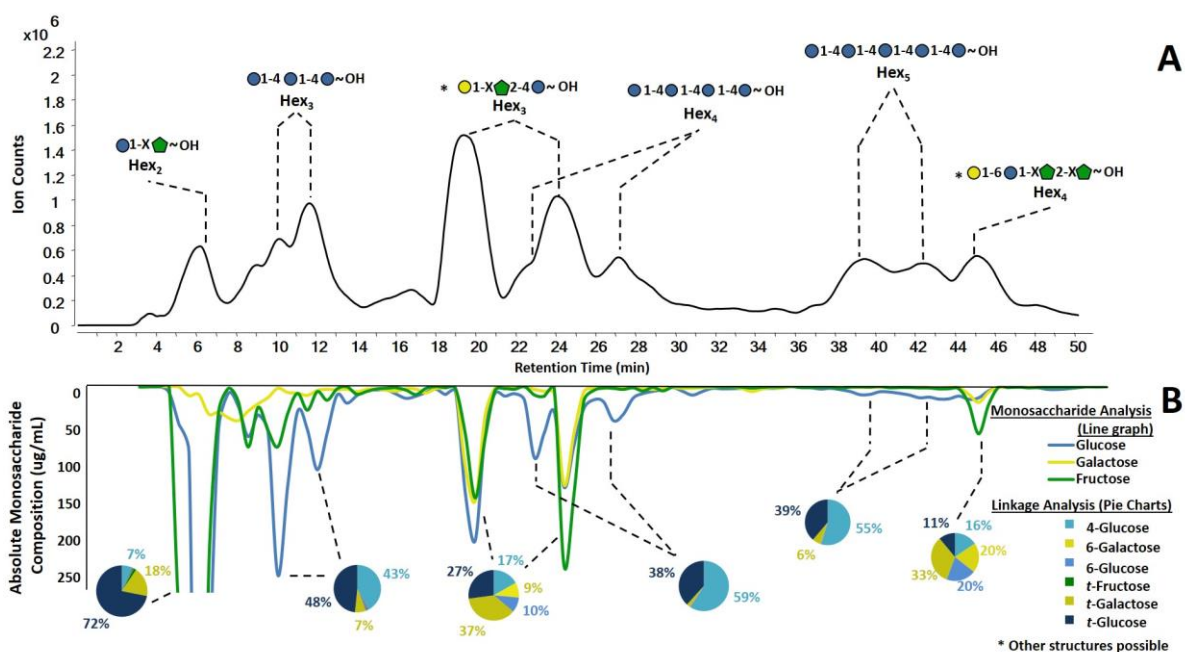
Figure 3.7. Results from the 3D LC-MS/MS workflow for butternut squash oligosaccharides.

(A) Oligosaccharide profiles from QTOF MS analysis with annotated elucidated structures. (B)

Mirrored chromatogram represents the absolute monosaccharide compositions of each fraction. Linkage compositions are shown in pie charts. In the inset structures, the "X" denote undefined linkages.

Oligosaccharides with mixed monosaccharide compositions were also present. A disaccharide was found to consist of glucose (49.5%) and fructose (50.5%), while the glycosidic linkage revealed the presence of t-glucose. Fructose is a ketosugar and does not derivatize well with PMP, therefore its linkages could not be determined. However, the results from the LC-QTOF MS showed the presence of hexose disaccharides. The structure was consistent with a disaccharide having a fructose reducing end and glucose nonreducing end (inset, **Figure 3.7A**). A trisaccharide was observed clustered at retention times 18-26 minutes. The oligosaccharide was composed of glucose (Glc, 35.6%), galactose (Gal, 26.8%), and fructose (Fru, 37.7%). Tandem MS yielded identical fragmentation. The multiple peaks suggested the presence of a reducing monosaccharide and the presence of anomers. Fructose does not produce anomers at the reducing end, eliminating the possibility of fructose being on the reducing end.[50,51] The linkage analysis of the oligosaccharide suggested the structure as Gal-(1→X)-Fru-(2→4)-Glc. The "X" corresponds to undefined linkage for the reasons written in the text below. Lastly, a tetrasaccharide at 45 minutes was found to consist of glucose (20.4%), galactose (20.9%), and fructose (58.8%). The linkage analysis revealed t-galactose (33.3%), and 6-glucose (20.1%). Because t-galactose (33.2 %) is higher in abundance compared to t-glucose (11.1 %), the t-galactose was assigned the terminal position leaving the 6-glucose to be in the internal position. The monosaccharide analysis revealed the presence of galactose and glucose residues, and linkage analysis yielded the structure Gal-(1→6)-Glc-(1→X)-Fru-(2→X)-Fru.

Table 3.4. Table of saccharides characterized using the *de novo* structural analysis of oligosaccharides generated from butternut squash with retention time, monosaccharide composition, and linkage composition.

Compound	*m/z* [M+H]⁺	RT (min)	Monosaccharide composition (%)			Linkage composition (%)					
			Glc	Gal	Fru	4-Glc	6-Gal	6-Glc	*t*-Fru	*t*-Gal	*t*-Glc
2Hex	343.116	6	43.8	5.1	51.1	6.7	0.6	0.6	1.8	18.5	71.9
3Hex	505.180	11	81.7	6.8	11.5	42.5	0.8	0.8	0.3	7.3	48.4
3Hex	505.180	19,24	35.6	26.8	37.7	16.8	9.6	9.6	0.1	36.8	27.1
4Hex	667.235	27	90.4	2.4	7.2	59.0	0.3	0.3	0.0	2.2	38.2
5Hex	829.285	41	78.1	7.9	14.1	54.8	0.2	0.2	0.0	6.0	38.8
4Hex	667.231	45	20.4	20.9	58.8	15.5	20.1	20.1	0.0	33.2	11.1

Glc- Glucose

Gal- Galactose

Fru- Fructose

Limitation of 3D MS for oligosaccharide analysis. The concept of 3D MS provides structural information of oligosaccharides that have previously not been achievable by any single method. However, there are clear limitations to the method. Coelution of several compounds can limit the specificity of the analysis. This condition can be remedied by further optimizing the separation conditions, by using a different stationary phase such as HILIC, and by significantly longer chromatographic separation.

The unique structure of fructose makes it difficult to obtain linkage information. Being a keto sugar, it was difficult to label with PMP, and its sensitivity was significantly lower compared to other monosaccharides. In the partially methylated form, the monosaccharide appeared even less reactive to PMP thereby restricting the formation of labeled, partially methylated fructose. The differences in reactivity need further investigation. However, the

terminal fructose was observable and measured. Work is currently being performed to find a better labeling method for fructose.

There are additional limitations with the method serving as a reminder of the difficulties in elucidating complete structures. The α/β character of the glycosidic linkages could not be resolved using the 3D MS workflow. However, the workflow can differentiate oligosaccharide isomers by yielding differential retention times for distinct α/β linkages. For example, maltotriose with α(1→4) linkages and cellotriose with β(1→4) linkages are isomers with widely different retention times (a difference of as much as five minutes was observed on the PGC column). Furthermore, other techniques may be used and adapted to the workflow. One approach is enzymatic treatments that can be applied on separated oligosaccharides.[52] Enzymes can be targeted based on the elucidated monosaccharide and linkage information.

CONCLUSIONS

A 3D LC-MS/MS-based workflow was developed for the structural elucidation of oligosaccharides derived from plant polysaccharides. This workflow can be employed to pure oligosaccharides, polysaccharides, and complicated polysaccharide mixtures such as those found in plants or food. This approach utilized LC-MS with inline collection of fractions that are further probed by separate LC-MS analyses yielding the corresponding monosaccharide compositions and linkage information on the eluting oligosaccharide. Method validation was performed on maltooligosaccharides, oligosaccharides derived from the polysaccharide galactan, and a food product, butternut squash. The optimized workflow yielded separated oligosaccharides with structures. The approach can also be applied to rapidly build glycome libraries of new and novel oligosaccharides.

REFERENCES

(1)   Engfer, M. B.; Stahl, B.; Finke, B.; Sawatzki, G.; Daniel, H. Human Milk Oligosaccharides Are Resistant to Enzymatic Hydrolysis in the Upper Gastrointestinal Tract. *Am. J. Clin. Nutr.* **2000**, *71* (6), 1589–1596. https://doi.org/10.1093/ajcn/71.6.1589.

(2)   Garrido, D.; Kim, J. H.; German, J. B.; Raybould, H. E.; Mills, D. A. Oligosaccharide Binding Proteins from *Bifidobacterium Longum* Subsp. Infantis Reveal a Preference for Host Glycans. *PLoS One* **2011**, *6* (3), e17315. https://doi.org/10.1371/journal.pone.0017315.

(3)   Marcobal, A.; Barboza, M.; Froehlich, J. W.; Block, D. E.; German, J. B.; Lebrilla, C. B.; Mills, D. A. Consumption of Human Milk Oligosaccharides by Gut-Related Microbes. *J. Agric. Food Chem.* **2010**, *58* (9), 5334–5340. https://doi.org/10.1021/jf9044205.

(4)   Sela, D. A.; Garrido, D.; Lerno, L.; Wu, S.; Tan, K.; Eom, H. J.; Joachimiak, A.; Lebrilla, C. B.; Mills, D. A. *Bifidobacterium Longum* Subsp. Infantis ATCC 15697 α-Fucosidases Are Active on Fucosylated Human Milk Oligosaccharides. *Appl. Environ. Microbiol.* **2012**, *78* (3), 795–803. https://doi.org/10.1128/AEM.06762-11.

(5)   Goubet, F.; Jackson, P.; Deery, M. J.; Dupree, P. Polysaccharide Analysis Using Carbohydrate Gel Electrophoresis. A Method to Study Plant Cell Wall Polysaccharides and Polysaccharide Hydrolases. *Anal. Biochem.* **2002**, *300* (1), 53–63. https://doi.org/10.1006/abio.2001.5444.

(6)   Kobata, A. Structures and Application of Oligosaccharides in Human Milk. *Proc. Japan Acad. Ser. B Phys. Biol. Sci.* **2010**, *86* (7), 731–747. https://doi.org/10.2183/pjab.86.731.

(7)   Newburg, D. S.; Neubauer, S. H. Carbohydrates in Milks: Analysis, Quantities, and Significance. In *Handbook of Milk Composition*; 1995; pp 273–350. https://doi.org/10.1016/b978-012384430-9/50015-9.

(8)   Wu, S.; Tao, N.; German, J. B.; Grimm, R.; Lebrilla, C. B. Development of an Annotated Library of Neutral Human Milk Oligosaccharides. *J. Proteome Res.* **2010**, *9* (8), 4138–4151. https://doi.org/10.1021/pr100362f.

(9)   Wu, S.; Grimm, R.; German, J. B.; Lebrilla, C. B. Annotation and Structural Analysis of Sialylated Human Milk Oligosaccharides. *J. Proteome Res.* **2011**, *10* (2), 856–868. https://doi.org/10.1021/pr101006u.

(10)   Davis, J. C. C.; Totten, S. M.; Huang, J. O.; Nagshbandi, S.; Kirmiz, N.; Garrido, D. A.; Lewis, Z. T.; Wu, L. D.; Smilowitz, J. T.; German, J. B.; Mills, D. A.; Lebrilla, C. B. Identification of Oligosaccharides in Feces of Breast-Fed Infants and Their Correlation with the Gut Microbial Community. *Mol. Cell. Proteomics* **2016**, *15* (9), 2987–3002.

https://doi.org/10.1074/mcp.M116.060665.

(11)   Varki, A. Glycan-Based Interactions Involving Vertebrate Sialic-Acid-Recognizing Proteins. *Nature* **2007**, *446* (7139), 1023–1029. https://doi.org/10.1038/nature05816.

(12)   Yoo, S. W.; Motari, M. G.; Susuki, K.; Prendergast, J.; Mountney, A.; Hurtado, A.; Schnaar, R. L. Sialylation Regulates Brain Structure and Function. *FASEB J.* **2015**, *29* (7), 3040–3053. https://doi.org/10.1096/fj.15-270983.

(13)   Kleene, R.; Schachner, M. Glycans and Neural Cell Interactions. *Nat. Rev. Neurosci.* **2004**, *5* (3), 195–208. https://doi.org/10.1038/nrn1349.

(14)   Spillmann, D.; Burger, M. M. Carbohydrate-Carbohydrate Interactions in Adhesion. *J. Cell. Biochem.* **1996**, *61* (4), 562–568. https://doi.org/10.1002/(SICI)1097-4644(19960616)61:4<562::AID-JCB9>3.0.CO;2-M.

(15)   Etzold, S.; Kober, O. I.; Mackenzie, D. A.; Tailford, L. E.; Gunning, A. P.; Walshaw, J.; Hemmings, A. M.; Juge, N. Structural Basis for Adaptation of Lactobacilli to Gastrointestinal Mucus. *Environ. Microbiol.* **2014**, *16* (3), 888–903. https://doi.org/10.1111/1462-2920.12377.

(16)   Gehrig, J. L.; Venkatesh, S.; Chang, H. W.; Hibberd, M. C.; Kung, V. L.; Cheng, J.; Chen, R. Y.; Subramanian, S.; Cowardin, C. A.; Meier, M. F.; O'Donnell, D.; Talcott, M.; Spears, L. D.; Semenkovich, C. F.; Henrissat, B.; Giannone, R. J.; Hettich, R. L.; Ilkayeva, O.; Muehlbauer, M.; Newgard, C. B.; Sawyer, C.; Head, R. D.; Rodionov, D. A.; Arzamasov, A. A.; Leyn, S. A.; Osterman, A. L.; Hossain, M. I.; Islam, M.; Choudhury, N.; Sarker, S. A.; Huq, S.; Mahmud, I.; Mostafa, I.; Mahfuz, M.; Barratt, M. J.; Ahmed, T.; Gordon, J. I. Effects of Microbiota-Directed Foods in Gnotobiotic Animals and Undernourished Children. *Science.* **2019**, *365* (6449), eaau4732. https://doi.org/10.1126/science.aau4732.

(17)   Hamaker, B. R.; Tuncil, Y. E. A Perspective on the Complexity of Dietary Fiber Structures and Their Potential Effect on the Gut Microbiota. *J. Mol. Biol.* **2014**, *426* (23), 3838–3850. https://doi.org/10.1016/j.jmb.2014.07.028.

(18)   Barboza, M.; Sela, D. A.; Pirim, C.; LoCascio, R. G.; Freeman, S. L.; German, J. B.; Mills, D. A.; Lebrilla, C. B. Glycoprofiling Bifidobacterial Consumption of Galacto-Oligosaccharides by Mass Spectrometry Reveals Strain-Specific, Preferential Consumption of Glycans. *Appl. Environ. Microbiol.* **2009**, *75* (23), 7319–7325. https://doi.org/10.1128/AEM.00842-09.

(19)   Sabater-Molina, M.; Larqué, E.; Torrella, F.; Zamora, S. Dietary Fructooligosaccharides and Potential Benefits on Health. *J. Physiol. Biochem.* **2009**, *65* (3), 315–328. https://doi.org/10.1007/BF03180584.

(20)   Singh, S. P.; Jadaun, J. S.; Narnoliya, L. K.; Pandey, A. Prebiotic Oligosaccharides: Special Focus on Fructooligosaccharides, Its Biosynthesis and Bioactivity. *Appl. Biochem.*

Biotechnol. **2017**, *183* (2), 613–635. https://doi.org/10.1007/s12010-017-2605-2.

(21) Kaplan, H.; Hutkins, R. W. Fermentation of Fructooligosaccharides by Lactic Acid Bacteria and Bifidobacteria. *Appl. Environ. Microbiol.* **2000**, *66* (6), 2682–2684. https://doi.org/10.1128/AEM.66.6.2682-2684.2000.

(22) Fehlbaum, S.; Prudence, K.; Kieboom, J.; Heerikhuisen, M.; van den Broek, T.; Schuren, F. H. J.; Steinert, R. E.; Raederstorff, D. In Vitro Fermentation of Selected Prebiotics and Their Effects on the Composition and Activity of the Adult Gut Microbiota. *Int. J. Mol. Sci.* **2018**, *19* (10), 3097. https://doi.org/10.3390/ijms19103097.

(23) Ziegler, A.; Zaia, J. Size-Exclusion Chromatography of Heparin Oligosaccharides at High and Low Pressure. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **2006**, *837* (1–2), 76–86. https://doi.org/10.1016/j.jchromb.2006.04.013.

(24) Bao, Y.; Newburg, D. S. Capillary Electrophoresis of Acidic Oligosaccharides from Human Milk. *Electrophoresis* **2008**, *29* (12), 2508–2515. https://doi.org/10.1002/elps.200700873.

(25) Corzo-Martínez, M.; Copoví, P.; Olano, A.; Moreno, F. J.; Montilla, A. Synthesis of Prebiotic Carbohydrates Derived from Cheese Whey Permeate by a Combined Process of Isomerisation and Transgalactosylation. *J. Sci. Food Agric.* **2013**, *93* (7), 1591–1597. https://doi.org/10.1002/jsfa.5929.

(26) Füreder, V.; Rodriguez-Colinas, B.; Cervantes, F. V.; Fernandez-Arrojo, L.; Poveda, A.; Jimenez-Barbero, J.; Ballesteros, A. O.; Plou, F. J. Selective Synthesis of Galactooligosaccharides Containing β(1→3) Linkages with β-Galactosidase from *Bifidobacterium Bifidum* (Saphera). *J. Agric. Food Chem.* **2020**, *68* (17), 4930–4938. https://doi.org/10.1021/acs.jafc.0c00997.

(27) Benkeblia, N. Fructooligosaccharides and Fructans Analysis in Plants and Food Crops. *J. Chromatogr. A* **2013**, *1313*, 54–61. https://doi.org/10.1016/j.chroma.2013.08.013.

(28) Sabater, C.; Prodanov, M.; Olano, A.; Corzo, N.; Montilla, A. Quantification of Prebiotics in Commercial Infant Formulas. *Food Chem.* **2016**, *194*, 6–11. https://doi.org/10.1016/j.foodchem.2015.07.127.

(29) Montilla, A.; Van De Lagemaat, J.; Olano, A.; Del Castillo, M. D. Determination of Oligosaccharides by Conventional High-Resolution Gas Chromatography. *Chromatographia* **2006**, *63* (9–10), 453–458. https://doi.org/10.1365/s10337-006-0770-5.

(30) Biermann, C. J.; McGinnis, G. D. *Analysis of Carbohydrates by GLC and MS*; 1988.

(31) Ruiz-Matute, A. I.; Hernández-Hernández, O.; Rodríguez-Sánchez, S.; Sanz, M. L.; Martínez-Castro, I. Derivatization of Carbohydrates for GC and GC-MS Analyses. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **2011**, *879* (17–18), 1226–1240. https://doi.org/10.1016/j.jchromb.2010.11.013.

(32) Black, I.; Heiss, C.; Azadi, P. Comprehensive Monosaccharide Composition Analysis of Insoluble Polysaccharides by Permethylation to Produce Methyl Alditol Derivatives for Gas Chromatography/Mass Spectrometry. *Anal. Chem.* **2019**, *91* (21), 13787–13793. https://doi.org/10.1021/acs.analchem.9b03239.

(33) Hsu, H. C.; Liew, C. Y.; Huang, S. P.; Tsai, S. T.; Ni, C. K. Simple Approach for De Novo Structural Identification of Mannose Trisaccharides. *J. Am. Soc. Mass Spectrom.* **2018**, *45* (3), 470–480. https://doi.org/10.1007/s13361-017-1850-5.

(34) Hsu, H. C.; Liew, C. Y.; Huang, S. P.; Tsai, S. T.; Ni, C. K. Simple Method for de Novo Structural Determination of Underivatised Glucose Oligosaccharides. *Sci. Rep.* **2018**, *8* (1), 1–12. https://doi.org/10.1038/s41598-018-23903-4.

(35) Cancilla, M. T.; Wong, A. W.; Voss, L. R.; Lebrilla, C. B. Fragmentation Reactions in the Mass Spectrometry Analysis of Neutral Oligosaccharides. *Anal. Chem.* **1999**, *71* (15), 3206–3218. https://doi.org/10.1021/ac9813484.

(36) Fasciotti, M.; Sanvido, G. B.; Santos, V. G.; Lalli, P. M.; McCullagh, M.; De Sá, G. F.; Daroda, R. J.; Peter, M. G.; Eberlin, M. N. Separation of Isomeric Disaccharides by Traveling Wave Ion Mobility Mass Spectrometry Using CO_2 as Drift Gas. *J. Mass Spectrom.* **2012**, *47* (12), 1643–1647. https://doi.org/10.1002/jms.3089.

(37) Hofmann, J.; Hahm, H. S.; Seeberger, P. H.; Pagel, K. Identification of Carbohydrate Anomers Using Ion Mobility-Mass Spectrometry. *Nature* **2015**, *526* (7572), 241–244. https://doi.org/10.1038/nature15388.

(38) Gray, C. J.; Thomas, B.; Upton, R.; Migas, L. G.; Eyers, C. E.; Barran, P. E.; Flitsch, S. L. Applications of Ion Mobility Mass Spectrometry for High Throughput, High Resolution Glycan Analysis. *Biochim. Biophys. Acta - Gen. Subj.* **2016**, *1860* (8), 1688–1709. https://doi.org/10.1016/j.bbagen.2016.02.003.

(39) Amicucci, M. J.; Nandita, E.; Lebrilla, C. B. Function without Structures: The Need for In-Depth Analysis of Dietary Carbohydrates. *J. Agric. Food Chem.* **2019**, *67* (16), 4418–4424.

(40) Amicucci, M. J.; Galermo, A. G.; Guerrero, A.; Treves, G.; Nandita, E.; Kailemia, M. J.; Higdon, S. M.; Pozzo, T.; Labavitch, J. M.; Bennett, A. B. A Strategy for Structural Elucidation of Polysaccharides: Elucidation of a Maize Mucilage That Harbors Diazotrophic Bacteria. *Anal. Chem.* **2019**, *54* (20), 7471–7480.

(41) Galermo, A. G.; Nandita, E.; Barboza, M.; Amicucci, M. J.; Vo, T.-T. T.; Lebrilla, C. B. Liquid Chromatography–Tandem Mass Spectrometry Approach for Determining Glycosidic Linkages. *Anal. Chem.* **2018**, *90* (21), 13073–13080.

(42) Galermo, A. G.; Nandita, E.; Castillo, J. J.; Amicucci, M. J.; Lebrilla, C. B. Development of an Extensive Linkage Library for Characterization of Carbohydrates. *Anal. Chem.* **2019**, *91* (20), 13022–13031.

(43)   Ninonuevo, M. R.; Park, Y.; Yin, H.; Zhang, J.; Ward, R. E.; Clowers, B. H.; German, J. B.; Freeman, S. L.; Killeen, K.; Grimm, R.; Lebrilla, C. B. A Strategy for Annotating the Human Milk Glycome. *J. Agric. Food Chem.* **2006**, *54* (20), 7471–7480. https://doi.org/10.1021/jf0615810.

(44)   Ruhaak, L. R.; Deelder, A. M.; Wuhrer, M. Oligosaccharide Analysis by Graphitized Carbon Liquid Chromatography-Mass Spectrometry. *Anal. Bioanal. Chem.* **2009**, *394* (1), 163–174. https://doi.org/10.1007/s00216-009-2664-5.

(45)   Kailemia, M. J.; Ruhaak, L. R.; Lebrilla, C. B.; Amster, I. J. Oligosaccharide Analysis by Mass Spectrometry: A Review of Recent Developments. *Anal. Chem.* **2014**, *86* (1), 196–212. https://doi.org/10.1021/ac403969n.

(46)   Lin, C. W.; Tsai, M. H.; Li, S. T.; Tsai, T. I.; Chu, K. C.; Liu, Y. C.; Lai, M. Y.; Wu, C. Y.; Tseng, Y. C.; Shivatare, S. S.; Wang, C. H.; Chao, P.; Wang, S. Y.; Shih, H. W.; Zeng, Y. F.; You, T. H.; Liao, J. Y.; Tu, Y. C.; Lin, Y. S.; Chuang, H. Y.; Chen, C. L.; Tsai, C. S.; Huang, C. C.; Lin, N. H.; Ma, C.; Wu, C. Y.; Wong, C. H. A Common Glycan Structure on Immunoglobulin G for Enhancement of Effector Functions. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (34), 10611–10616. https://doi.org/10.1073/pnas.1513456112.

(47)   Amicucci, M. J.; Galermo, A. G.; Nandita, E.; Vo, T.-T. T.; Liu, Y.; Lee, M.; Xu, G.; Lebrilla, C. B. A Rapid-Throughput Adaptable Method for Determining the Monosaccharide Composition of Polysaccharides. *Int. J. Mass Spectrom.* **2019**, *438*, 22–28.

(48)   Xu, G.; Amicucci, M. J.; Cheng, Z.; Galermo, A. G.; Lebrilla, C. B. Revisiting Monosaccharide Analysis–Quantitation of a Comprehensive Set of Monosaccharides Using Dynamic Multiple Reaction Monitoring. *Analyst* **2018**, *143* (1), 200–207.

(49)   Amicucci, M. J.; Nandita, E.; Galermo, A. G.; Castillo, J. J.; Chen, S.; Park, D.; Smilowitz, J. T.; German, J. B.; Mills, D. A.; Lebrilla, C. B. A Nonenzymatic Method for Cleaving Polysaccharides to Yield Oligosaccharides for Structural Analysis. *Nat. Commun.* **2020**, *11* (1), 3963. https://doi.org/10.1038/s41467-020-17778-1.

(50)   Hogarth, A. J. C. L.; Hunter, D. E.; Jacobs, W. A.; Garleb, K. A.; Wolf, B. W. Ion Chromatographic Determination of Three Fructooligosaccharide Oligomers in Prepared and Preserved Foods. *J. Agric. Food Chem.* **2000**, *48* (11), 5326–5330. https://doi.org/10.1021/jf000111h.

(51)   Li, J.; Hu, D.; Zong, W.; Lv, G.; Zhao, J.; Li, S. Determination of Inulin-Type Fructooligosaccharides in Edible Plants by High-Performance Liquid Chromatography with Charged Aerosol Detector. *J. Agric. Food Chem.* **2014**, *62* (31), 7707–7713. https://doi.org/10.1021/jf502329n.

(52)   Xu, G.; Goonatilleke, E.; Wongkham, S.; Lebrilla, C. B. Deep Structural Analysis and Quantitation of O-Linked Glycans on Cell Membrane Reveal High Abundances and Distinct Glycomic Profiles Associated with Cell Type and Stages of Differentiation. *Anal.*

Chem. **2020**, *92* (5), 3758–3768. https://doi.org/10.1021/acs.analchem.9b05103.

Chapter 4

A Method for Monitoring Glycosidic Linkages in

Food and Feces for Clinical Trials

ABSTRACT

Carbohydrates are the largest components in plant-based foods and play important roles in many biological functions. Dietary carbohydrates that are non-digestible by the host and end up in the distal gut as food for gut bacteria. Glycosidic linkages play an important role in this interaction as bacterial enzymes have high linkage specificity to degrade the carbohydrate substrates. The structural diversity of dietary carbohydrate in addition to their complicated matrices, such as foods and feces, makes their analysis difficult. In this research, a high throughput linkage analysis method was established to monitor food-microbe interactions in clinical trials for mice and human studies. A total of 33 glycosidic linkages were measured from a library capable of detecting up to 96 linkages. The quantitative approach revealed differences in carbohydrate linkage abundances in germ-free mice feces, and microbiota inoculated feces from mice. Altered amounts of 5-arabinose and T-arabinose in fecal samples from human subjects that were fed pea fiber snack were observed, indicating utilization of arabinans by gut microbes. The ultra-high performance liquid chromatography-triple quadrupole mass spectrometry (UHPLC-QqQ MS) method can lessen the time for quantitating glycosidic linkages in clinical trials and is useful tool to monitor dietary carbohydrate degradation in the presence of microbial communities.

INTRODUCTION

Carbohydrates represent the majority composition in plant-based foods. Food carbohydrates are composed of a variety of different molecular structures such as monosaccharides, disaccharides, oligosaccharides, and polysaccharides. These biomolecules often have associated functions that are dictated by their specific structures.[1] For example, starch polysaccharides, a glucose polymer with a ($\alpha1\rightarrow4$) glycosidic linkage for amylose and amylopectin also contains a glucose-based polymer with a ($\alpha1\rightarrow4$) glycosidic linkage; however, it branches every 12-25 glucose units at ($\alpha1\rightarrow6$) glycosidic linkage positions.[2, 3] Starch digestion starts at the mouth where salivary amylases catalyze the hydrolysis of starch at the specific ($\alpha1\rightarrow4$) glycosidic linkage position.[4] Another function of cell wall polysaccharides such as cellulose, xyloglucan, pectins, hemicelluloses is to provide the plant cell with structural rigidity to withstand stresses.[5] However, these polysaccharides are non-digestible by the human host, and can be a food source for gut microbes.[6, 7] These organisms contain genes encoding for the production of carbohydrate-active enzymes (CAZymes) that degrade and utilize the selective carbohydrate structures.

In recent years, there has been great interest in the gut microbiome capabilities to affect host health. The interactions between dietary carbohydrates and the gut microbiome produces metabolites such as short chain fatty acids (SCFA) that are associated with reducing the risk of various metabolic disorders.[8, 9] Similarly, specific diets have shown to alter bacterial species abundances in humans.[10, 11] Gut microbes can also play a role in mediating obesity. A high fat induced diet in mice and the presence of *B. longum* species showed reduction of body weight and fat accretion.[12] Thus, there is a need to characterize these interactions to mediate host health

outcomes. However, the structure to function relationships between dietary carbohydrates and microbial communities largely remains undetermined.

Glycosidic linkages are important features of carbohydrate structures. They alone can even be used to deduce parent polysaccharides structures. They can also reveal the specific substrate for glycosyl hydrolyses (GH) and polysaccharide lyases (PL) needed to degrade polymers. While methods for determining the CAZymes using genomic and GH databases were developed, the analysis of carbohydrate linkages has not advanced particularly compared to other omics methods such as genomic and proteomics.[13, 14] Traditional methods for linkage analysis are still performed on a GC-MS platform. The analysis requires extensive derivatization steps to make analytes amendable for the analysis.[15] Other limitations include: only a small set of linkages can be monitored, long chromatographic run times, and the general lack of sensitivity.[16, 17] Moreover, samples are prepared and analyzed on a single sample manner. This lowers the sample throughput and inhibits the processing of large clinical feeding studies samples. There remains a need to rapidly analyze carbohydrate linkages in a high-throughput manner.

Modern analytical approaches for the characterization of biological compounds employ liquid chromatography coupled to mass spectrometry (LC-MS). We recently published the first LC-MS method for the determination of glycosidic linkage that yields greater sensitivity and that is capable of monitoring many more linkages than previous approaches. It also includes faster chromatographic run times, thereby enhancing the speed of analysis.[18] Subsequently, an extended linkage library was developed that is capable of monitoring up to 96 different glycosidic linkages on the LC-MS platform.[19] The LC-MS approach provides better figures of merit compared to the GC-MS platform and is suitable for characterizing carbohydrates in large

sample batches. Currently, linkage analysis using a 96-well plate format for clinical feeding studies has not been reported.

In this work, an LC-MS method for the determination of glycosidic linkages was employed to a large feeding study to measure the linkages that were consumed by bacteria in the gut. Fibers, mice diets, mice fecal and human fecal samples were subjected to the high-throughput glycosidic linkage analysis. Samples were subjected to permethylation, hydrolysis, and derivatization by PMP subsequently analyzed by UHPLC-QqQ MS. The method was performed on over 200 fecal and fiber samples. Technical replicates (n=3) for each sample were performed using the linkage analysis 96 well plate format to demonstrate the methods robustness. The sensitivity allowed the large sample set to be processed with only 50 µg of feces per analysis. The measured glycosidic linkages abundances were then compared to evaluate the utilization of carbohydrates structures by gut microbes.

EXPERIMENTAL PROCEDURES

Samples and materials. Fibers, diets, and fecal biospecimens samples were obtained from a collaborative clinically feeding study.[20] Iodomethane (contains copper stabilizer, 99.5%), trifluoroacetic acid (TFA) (HPLC grade), chloroform (HPLC grade), ammonium acetate (NH4AC), sodium hydroxide pellets (semi-conductor grade 99.99% trace metals basis), ammonium hydroxide solution (NH4OH) (28-30%), dichloromethane (DCM), 3-methyl-1-phenyl-2-pyrazoline-5-one (PMP), methanol (HPLC grade), and anhydrous dimethyl sulfoxide (DMSO) were purchased from Sigma-Aldrich (St. Louis, MO). Amylopectin, 1,6-α-D-mannotriose, maltohexaose, 1,4-β-D-mannotriose, 3-O-(α-D-mannopyranosyl)-D-mannopyranose, 3-O-(β-D-galactopyranosyl)-D-galactopyranose, nigerose, lactose, 2′-fucosyllactose (synthetic), 4-O-(β-D-

galactopyranosyl)-D-galactopyranose, isomaltotriose, 1,3-α-1,6-α-D-mannotriose, 1,4-D-xylobiose, 1,5-α-Larabinotriose, and 2-O-(α-D-Mannopyranosyl)-D-mannopyranose were purchased from Carbosnth (Compton, U.K.). Arabinoxylan was purchased from Megazyme (Chicago, IL). Acetonitrile (HPLC grade) was purchased from Honeywell (Muskegon, MI). Nanopure water was used for all experiments.

Sample preparation and homogenization. Fibers, mice diets, and fecal biospecimens were treated to complete dryness by lyophilization. Dried samples underwent a homogenization step by grinding to a fine powder using a bead mill homogenizer. A representative 10 mg aliquot was taken to make a final stock solution of 10 mg/mL. Stock solutions underwent a homogenization step by bead milling using 2 mm stainless steel beads. Samples were then incubated at 100°C for 1h to aid in solubilizing polysaccharides. Lastly, another round of bead milling was employed to complete the homogenization process. Aliquots from homogenized stock solution were used for linkage analysis.

Linkage analysis by LC-MS. Linkage analysis was adapted from Galermo et al. with the following modifications.[18] In short, three 5 μL replicates from each stock solution from feeding study and a pool of oligosaccharide standards used as quality controls were transferred onto a 96-well plate. Samples reacted in a mixture containing saturated NaOH and iodomethane in DMSO. Residual DMSO and NaOH were removed by extraction with water and DCM. The DCM layer was completely dried using vacuum centrifugation. Samples were hydrolyzed using 4 M TFA for 2 h at 100°C. The resulting products were subjected to PMP derivatization in 0.2 M PMP in methanol and 28% NH_4OH for 30 min at 70˚C. The derivatized compounds were then dried to complete by vacuum centrifugation. The labeled methylated species were reconstituted in 70%

methanol/water (v/v) prior to LC-MS analysis. For each sample, a 1 µL aliquot was analyzed on an Agilent 1290 infinity II UHPLC coupled to an Agilent 6495A QqQ MS operated in multiple reaction monitoring (MRM) mode. A pool of oligosaccharide standards was used to assign retention times and identify glycosidic linkages in the samples. Raw LC-MS files were analyzed using Agilent MassHunter Quantitative Analysis software (Version B 08.00). The MRM peaks were manually integrated using area of the curve, for each linkage.

RESULTS AND DISCUSSION

A comprehensive LC-MS glycosidic linkage analysis was employed on samples from feeding studies comprised of food fibers, diets, feces from mice and humans. Quantitation of glycosidic linkages in these complicated matrices were measured using a targeted tandem mass spectrometry approach employing a multiple reaction monitoring (MRM) method. For each linkage, two MRM transitions corresponding to a quantifier and qualifier product ions were used. The feeding study samples were subjected to the workflow shown in **Figure 4.1**. First, the process entailed a lyophilization step to remove moisture and normalize the quantitation of linkages measured by dry weight of the samples. Next, homogenization steps were employed to ensure linkages from insoluble and soluble polysaccharides were accurately accounted for and determined. The samples were then permethylated in a 96-well plate format followed by an acid hydrolysis treatment. The remaining products were subjected to PMP derivatization followed by UHPLC-QqQ MS analysis. This approach used a targeted MRM mode and a linkage library based on retention time and mass transitions to determine linkage profiles of fibers, diets, and feces. Finally, the measured linkage abundances were compared across the feeding study samples to reveal the carbohydrate utilization by gut microbes.
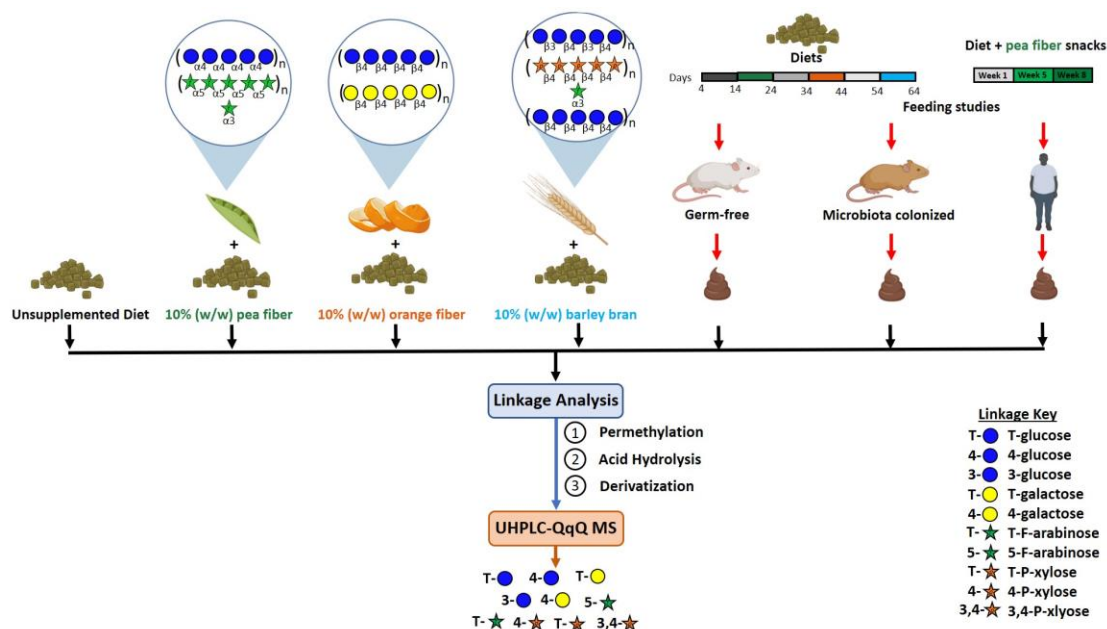
Figure 4.1. Workflow for the UHPLC-QqQ MS linkage analysis of fibers, diets, and feces.

Linkage composition of dietary fibers in mice diets

The fibers and mice diets were subjected to glycosidic linkage analysis. Three dietary fibers were chosen to supplement a base line comprised of a high saturated fat low fruits and vegetables (HiSF-LoFV) by 10 % (w/w). First, the glycosidic linkages were determined for the three dietary fibers: pea fiber, orange fiber, and barley bran. The results shown in **Figure 4.2A** illustrate linkages corresponding to the three dietary fibers and yielded over 28 linkages with a relative abundance ≥ 0.1 % listed in **Table 4.1**. The standard deviation in linkages detected were found to be ≤ 1.9 % demonstrating the robustness and experimental reproducibility for the linkage analysis of fibers.

In pea fiber, a high abundance of T-arabinose (22.1%) and 5-arabinose (16.5%) were observed. These linkages are common in arabinan and typically contain α(1→5)-arabinose backbones with capped T-arabinose at branched points α(1→2) and α(1→3) positions.[21, 22] The

97

lower composition of 3-arabinose (0.5%) and 2-arabinose (0.6%) suggested arabinan fiber to have a linear rather than branched structure.

In orange fiber, high amounts of T-arabinose (20.3 %), and 5-arabinose (10.3 %) were detected, indicating the presence of arabinan. A greater abundance in 3-arabinose (1.3 %) and 2-arabinose (1.4 %) in orange fiber were observed compared to pea fiber and suggested a more branched arabinan structure. It is worth noting that the same polysaccharides from different fiber sources may have slightly different structures. T-galactose (5.8 %) and 4-galactose (4.2 %), were present in orange fiber, and the combination of these linkages indicates the presence of galactan. Galactan from pectins contains repeating linear $\beta(1\rightarrow4)$-galactose units.[23] Furthermore, the detection of T-rhamnose (3.7 %) and T-Gal (0.1%) suggested the presence of pectin. Fruits, in particular citrus, have been shown to be composed of pectins containing both arabinan and galactan.[24]

Challenges arise for the linkage analysis of acid polysaccharides from pectins. The backbone of pectins typically contain galacturonic acid residues linked $\alpha(1\rightarrow4)$ and a carboxylic acid group at C6 position.[25] In general, these polymers are resistant to chemical hydrolysis making it difficult to produce liberated monosaccharides.[26] Pectin polysaccharides include a diverse set of polymers and can vary from different sources to include rhamnogalacturonan I (RGI), and rhamnogalacturonan II (RGII).[27] In future work, an optimized LC-MS/MS workflow will be developed to detect linkages from acid polymers. Despite this limitation, other linkages from pectin polymers can be used as markers for the identification of pectins. These markers include the presence of galactose, arabinose, galacturonic acid, rhamnose monosaccharides and linkages.

In barley bran, higher amounts of 3-glucose (12.1%) and 4-glucose (37.2 %) were found compared to those in pea and orange fiber. These linkages are found in β-glucan and are commonly present in barley brans.[28] Additionally, barley bran had a high composition of T-arabinose (14.2 %) and minimal linear arabinose linkages (internal residues) indicating the absences of arabinan. However, the combination of 4-xylose (3.2 %) and T-arabinose (14.2 %) suggested the presence of arabinoxylan. Brans in general are known to contain arabinoxylans.[29, 30] The relative linkage compositions of the three fibers are shown in **Figure 4.2A.**
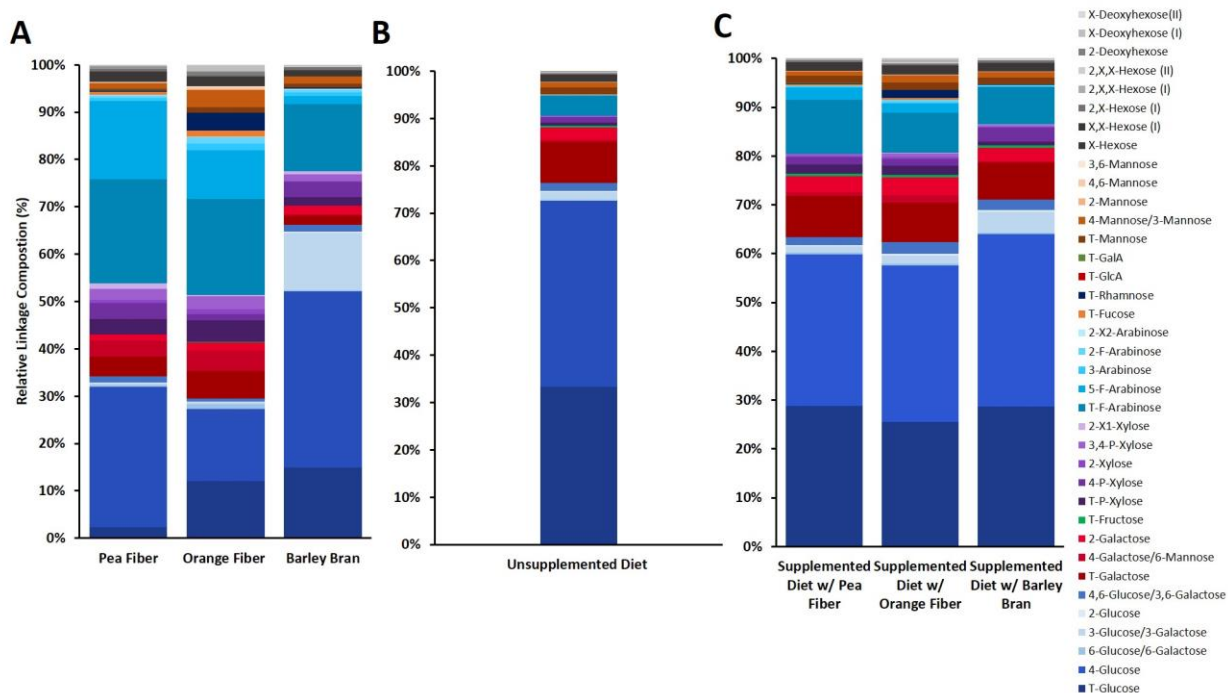

Figure 4.2. The relative linkage composition of (A) fibers, (B) unsupplemented base diet, and (C) supplemented base diets with 10 % (w/w) fiber.

The linkages from the baseline diet that is composed of high saturated fats and low fruits and vegetables without fiber supplementation were next determined. The linkage profile for this diet is displayed in **Figure 4.2B**. The results yielded 4-glucose (39.3 %) as the most abundant linkage followed by T-glucose (33.3 %). The prevalence of T-glucose and 4-glucose suggested starch in the form of amylose and amylopectin as the main polysaccharides. Amylose and amylopectin have a high abundance in α(1→4) glucose linkages and are a major energy source for the host metabolism of mice and humans.[31] The T-glucose (33.3 %) observed in great abundance and is partly coming from sucrose, where the glucose residue is represented at the terminal linkage species in the disaccharide. Sucrose is added in mice chow diets for preferential consumption.[32] In addition to the starch and sucrose linkages, T-galactose (8.7 %) and T-arabinose (4.2 %) were the next most abundant linkages detected. However, low relative abundances for linear linkages were observed, so it is difficult to determine the polysaccharide structures. Nonetheless, the carbohydrate linkages measured mainly corresponded to starch and sucrose.

Table 4.1. Relative linkage composition in the unsupplemented base diet and supplemented base diet with 10 % (w/w) of fibers. The deviation corresponded to the standard deviations of experimental replicates (n=3).

Linkage	HiSF-LoFV diets			
	Base	Base + pea fiber	Base + orange fiber	Base + barley bran
T-Glucose	33.2 ± 1.2	28.9 ± 0.6	25.7 ± 1.5	28.7 ± 1.4
4-Glucose	39.2 ± 1.1	31.1 ± 0.6	31.7 ± 3.3	35.1 ± 0.6
6-Glucose/6-Galactose	0.3 ± 0.0	0.3 ± 0.0	0.5 ± 0.1	0.3 ± 0.0
3-Glucose/3-Galactose	1.5 ± 0.1	1.2 ± 0.1	1.7 ± 0.9	4.3 ± 0.1
2-Glucose	0.3 ± 0.0	0.3 ± 0.0	0.4 ± 0.2	0.3 ± 0.0
4,6-Glucose/3,6-Galactose	1.8 ± 0.3	1.7 ± 0.2	2.4 ± 0.6	2.2 ± 0.4
T-Galactose	8.8 ± 1.0	8.4 ± 0.3	8.3 ± 1.5	7.6 ± 0.8

4-Galactose/6-Mannose	0.3 ± 0.0	0.8 ± 0.0	1.4 ± 0.1	0.2 ± 0.0
2-Galactose	2.6 ± 0.2	3.2 ± 1.0	3.6 ± 1.0	2.8 ± 0.0
T-Fructose	0.4 ± 0.1	0.5 ± 0.0	0.5 ± 0.2	0.5 ± 0.1
T-P-Xylose	0.7 ± 0.0	2.0 ± 0.1	1.9 ± 0.2	0.7 ± 0.1
4-P-Xylose	1.1 ± 0.2	1.5 ± 0.1	1.4 ± 0.3	2.9 ± 0.3
2-Xylose	0.1 ± 0.0	0.2 ± 0.0	0.4 ± 0.0	0.0 ± 0.0
3,4-P-Xylose	0.2 ± 0.0	0.4 ± 0.1	0.8 ± 0.0	0.4 ± 0.0
$2\text{-}X_1\text{-Xylose}$	0.0 ± 0.0	0.1 ± 0.0	0.1 ± 0.0	0.1 ± 0.1
T-F-Arabinose	4.2 ± 0.3	11 ± 0.1	8.2 ± 0.9	7.6 ± 1.2
5-F-Arabinose	0.2 ± 0.0	2.6 ± 0.1	1.9 ± 0.2	0.2 ± 0.0
3-F-Arabinose	0.1 ± 0.0	0.1 ± 0.0	0.3 ± 0.0	0.2 ± 0.0
2-F-Arabinose	0.1 ± 0.0	0.2 ± 0.0	0.5 ± 0.1	0.2 ± 0.0
$2\text{-}X_2\text{-Arabinose}$	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
T-Fucose	0.1 ± 0.0	0.3 ± 0.0	0.4 ± 0.0	0.1 ± 0.0
T-Rhamnose	0.0 ± 0.0	0.2 ± 0.0	1.6 ± 0.3	0.0 ± 0.0
T-GlcA	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
T-GalA	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
T-Mannose	1.5 ± 0.2	1.5 ± 0.1	1.5 ± 0.1	1.5 ± 0.1
4-Mannose/3-Mannose	1.0 ± 0.0	1.0 ± 0.0	1.5 ± 0.1	1.1 ± 0.0
2-Mannose	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
4,6-Mannose	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
3,6-Mannose	0.0 ± 0.0	0.0 ± 0.0	0.2 ± 0.0	0.0 ± 0.0
X-Hexose	0.1 ± 0.0	0.3 ± 0.0	0.3 ± 0.1	0.1 ± 0.0
X,X-Hexose (I)	1.4 ± 0.1	1.5 ± 0.1	1.6 ± 0.1	1.7 ± 0.1
2,X-Hexose (I)	0.3 ± 0.1	0.4 ± 0.0	0.4 ± 0.2	0.5 ± 0.1
2,X,X-Hexose (I)	0.0 ± 0.0	0.0 ± 0.0	0.1 ± 0.0	0.0 ± 0.0
2,X,X-Hexose (II)	0.2 ± 0.2	0.1 ± 0.0	0.4 ± 0.3	0.2 ± 0.1
2-Deoxyhexose	0.0 ± 0.0	0.1 ± 0.0	0.1 ± 0.0	0.0 ± 0.0
X-Deoxyhexose (I)	0.1 ± 0.0	0.1 ± 0.0	0.5 ± 0.1	0.1 ± 0.0
X-Deoxyhexose (II)	0.0 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0

To demonstrate changes in linkage profiles with the addition of a 10% (w/w) dietary fiber, the unsupplemented diet and supplemented diets were subjected to linkage analysis. A total of 37 linkage profiles were observed (**Table 4.1**). Overall, the supplemented base diet with a 10% (w/w) with fiber increased the relative fiber linkage abundances when compared to the unsupplemented base diet (**Figure 4.3**). These diets were then used in a mice feeding study

described below. The linkage detected are representative of the polysaccharides present in the samples and were monitored throughout the feeding study.

In the pea fiber supplemented base diet, a statistically significant increase in abundance of T-arabinose and 5-arabinose (10.1% and 2.6 %) were observed compared to the unsupplemented baseline diet (4.2% and 0.2%). The increased composition of T-arabinose and 5-arabinose indicated the addition of linear arabinan to the baseline diet. In the orange fiber supplemented base diet, an increase in 4-galactose abundance was observed. This increase of 4-galactose corresponds to an increase of galactan. A greater abundance of 3-glucose (4.3%) was observed in the barley bran supplemented diet compared to the unsupplemented diet. As expected, 3-glucose is present in β-glucan and is commonly found in brans. Furthermore, 4-xylose increased from 1.1 % to 2.9%, and is an indication of arabinoxylan addition to the diet. Overall, a 10 % (w/w) addition of fiber supplemented to the base diet increased the fiber linkage abundances (**Figure 4.3**).
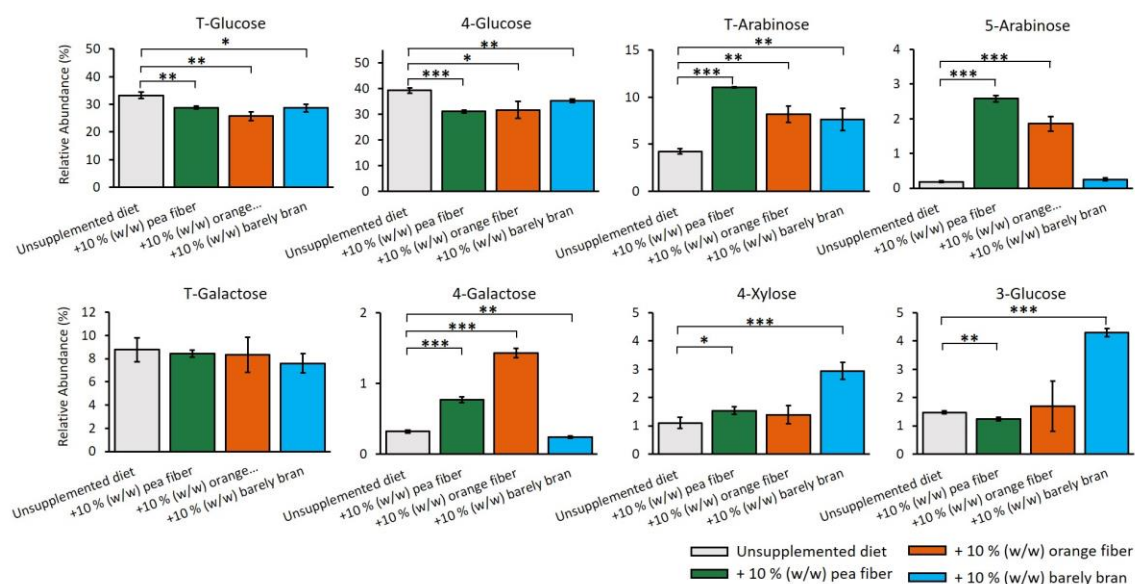
Figure 4.3. Linkage profiles in the unsupplemented base diet and supplemented base diets with 10 % (w/w) of fiber (n=3). Mean values ± SD were plotted, and error bars indicate to the standard deviations of 3 technical replicates. P-values correspond to * p>0.05 by t-test compared to the unsupplemented diet group.

Linkage analysis in mice feces.

To quantitate and monitor linkage abundances in feeding studies, the in-depth linkage analysis of the before (fibers and diets) and after (feces) *in vivo* digestion is needed. The linkage analysis of fibers and diets were described in the text above. Here, we exemplify the linkages analysis in a mice feeding study. The study included a 64-day feeding intervention with gnotobiotic mice colonized with obese human gut microbiota on a 10-day diet rotation. To control the food-microbial interactions, germ-free (GF) mice fecal samples were used as a control. Both microbiota colonized and germ-free mice were on a 'unhealthy' HiSF-LoFV diet representing the western diet throughout the study. The glycosidic linkages remaining in feces provides an overview of carbohydrates post-consumption by gut bacteria. The results yielded CVs < 20% for all fecal samples, illustrating the reproducibility of the method. Furthermore, statistically significant values were obtained and were used to compare linkage abundances across the feeding study.
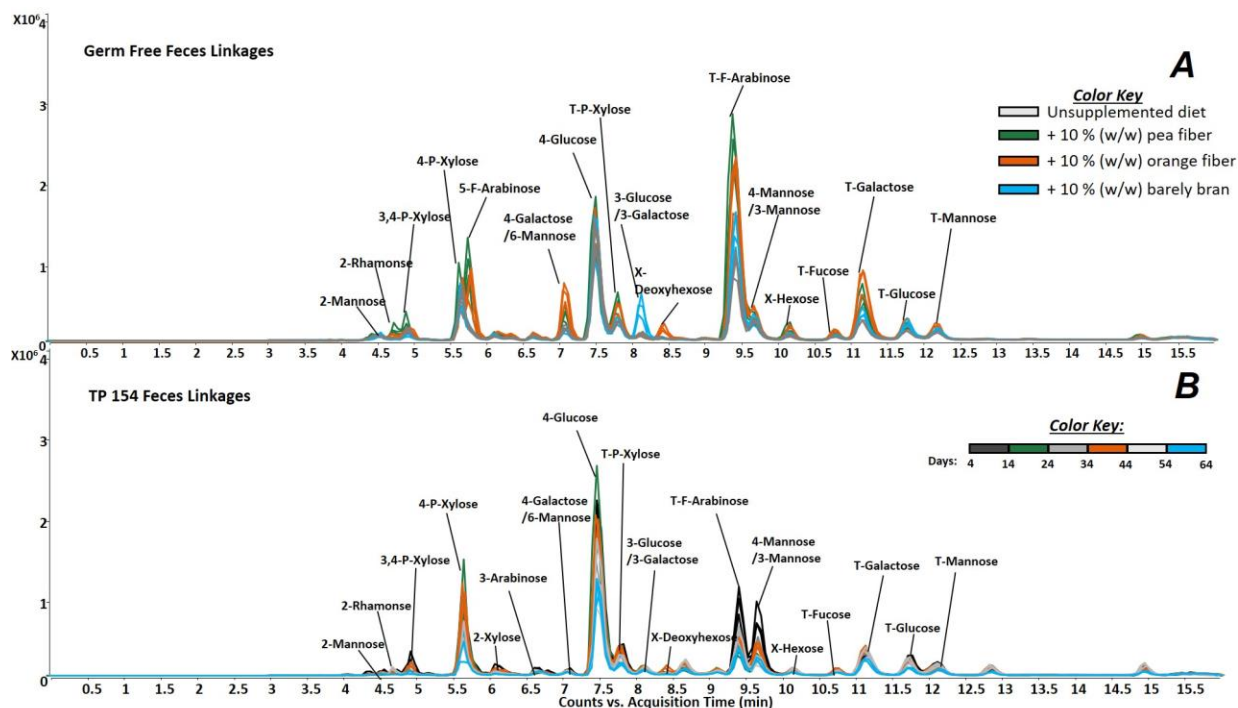
Figure 4.4. Representative MRM chromatogram for the linkage analysis of feces (A) germ-free and (B) microbiota colonized mice feces shown in three experimental replicates. The solid-colored lines represent the linkage profiles of the unsupplemented and supplemented baseline diets with 10 % (w/w) fiber (A). For (B), the solid-colored lines represent the linkages remaining in feces from the 64-day mice feeding trial.

The remaining carbohydrates in the GF mice feces corresponded to the non-digestible carbohydrates by host or dietary fibers. The LC chromatogram in **Figure 4.4A** revealed the linkages associated with polysaccharides remaining in the GF mice feces. For example, T-arabinose and 5-arabinose corresponding to arabinan were readily observed in the GF mice feces that were fed with the pea and orange fibers. For GF mice that were fed the barley bran diet, 3-glucose was solely observed. On the contrary, linkages associated with fibers in the 64-day mice feeding trial (colonized with gut microbes) were not observed for some and were minimally

present for others (**Figure 4.4B**). The linkages in the 64-day feeding study suggested the fiber

were utilized by the gut microbes. For example, in the pea and oranges fibers fed diets, the 5-

arabinose was completely consumed. Similarly, the 3-glucose present in barley bran was

completely consumed as well. A total of 39 glycosidic linkages were measured (**Figure 4.5A**)

germ free and (**Figure 4.5B**) microbe colonized feces.

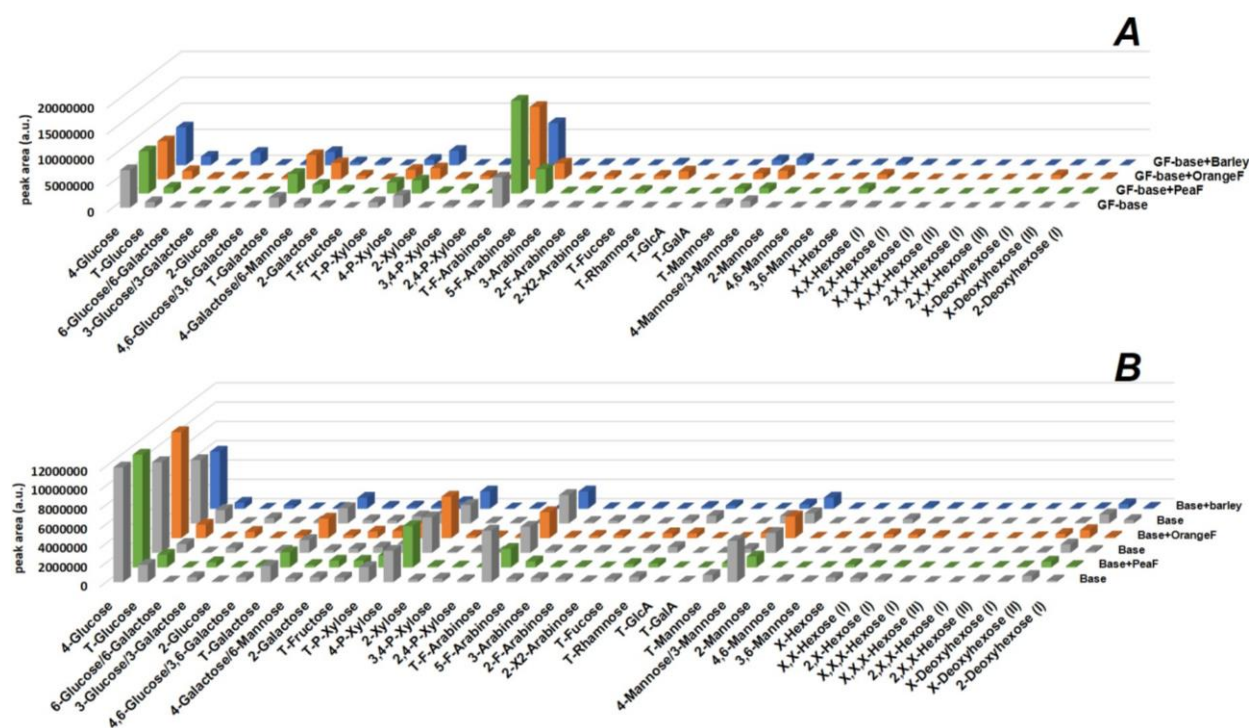**Figure 4.5.** Linkage profiles in mice feces samples from germ free feed with and without 10 %

(w/w) fiber supplements (A). Linkage distribution in a 10-day feeding study with base line and

fiber supplemented diets (B).

Table 4.2. List of relative linkages abundances in germ free and microbe colonized mice feces.

Relative linkage composition (%)	
Germ free mice feces	Microbe colonized mice feces

Linkage	GF base	GF base + pea fiber	GF base + orange fiber	GF base + barley bran	Base	Base + pea fiber	Base	Base + orange fiber	Base	Base + barley bran
T-Glucose	3.9 ± 0.8	2.3 ± 0.1	3.2 ± 0.3	5.1 ± 0.4	4.7 ± 0.9	4.5 ± 0.7	3.3 ± 0.3	4.1 ± 0.3	5.8 ± 0.4	3.6 ± 0.0
4-Glucose	26.3 ± 1.8	16 ± 0.1	14.6 ± 0.9	21.9 ± 1.9	30.8 ± 3.3	38.4 ± 2.1	33.6 ± 1.0	32.3 ± 0.9	26.7 ± 2.2	34 ± 1.5
6-Glucose/6-Galactose	0.5 ± 0.1	0.4 ± 0.0	0.7 ± 0.0	0.5 ± 0.0	0.2 ± 0.0	0.1 ± 0.0	0.2 ± 0.0	0.2 ± 0.0	0.2 ± 0.0	0.1 ± 0.0
3-Glucose/3-Galactose	1.4 ± 0.2	0.6 ± 0.0	0.8 ± 0.1	7.3 ± 0.6	1.3 ± 0.2	1.8 ± 0.1	1.8 ± 0.1	1.9 ± 0.1	2.2 ± 0.2	2.1 ± 0.1
2-Glucose	0.3 ± 0.0	0.2 ± 0.0	0.3 ± 0.0	0.4 ± 0.0	0.2 ± 0.0	0.2 ± 0.0	0.2 ± 0.0	0.2 ± 0.0	0.4 ± 0.0	0.3 ± 0.0
4,6-Glucose/3,6-Galactose	0.9 ± 0.1	0.5 ± 0.1	0.7 ± 0.1	0.9 ± 0.1	1.3 ± 0.5	0.7 ± 0.2	0.8 ± 0.2	0.8 ± 0.3	0.9 ± 0.1	0.6 ± 0.1
T-Galactose	7.1 ± 0.3	7.5 ± 0.1	9.2 ± 0.2	7.8 ± 0.5	4.4 ± 0.2	5.3 ± 0.7	4.7 ± 0.1	5.9 ± 0.2	6.6 ± 0.3	6.6 ± 0.3
4-Galactose/6-Mannose	2.9 ± 0.4	3.3 ± 0.2	6.1 ± 0.5	1.7 ± 0.1	1.0 ± 0.1	0.9 ± 0.0	1.1 ± 0.0	1.0 ± 0.1	1.4 ± 0.1	1.2 ± 0.0
2-Galactose	1.4 ± 0.1	1.2 ± 0.1	1.5 ± 0.1	1.2 ± 0.2	1.3 ± 0.1	2.2 ± 0.6	1.5 ± 0.5	2.0 ± 0.3	1.3 ± 0.1	1.4 ± 0.4
T-Fructose	0.2 ± 0.0	0.1 ± 0.0	0.1 ± 0.0	0.1 ± 0.0	1.2 ± 0.1	2.4 ± 0.5	2.0 ± 0.1	2.2 ± 0.1	3.1 ± 0.3	1.5 ± 0.1
T-P-Xylose	4.0 ± 0.2	4.3 ± 0.3	3.7 ± 0.3	3.2 ± 0.2	4.0 ± 0.1	4.2 ± 0.5	3.3 ± 0.2	4.4 ± 0.4	4.2 ± 0.3	4.0 ± 0.3
4-P-Xylose	8.8 ± 1.3	5.0 ± 0.6	4.2 ± 0.4	8.5 ± 0.5	8.4 ± 0.4	14.3 ± 0.9	13 ± 0.8	12.7 ± 0.6	8.0 ± 0.7	10.2 ± 2.3
2-Xylose	0.4 ± 0.0	0.6 ± 0.1	0.7 ± 0.1	0.2 ± 0.0	0.7 ± 0.0	0.6 ± 0.0	0.4 ± 0.0	0.8 ± 0.0	0.5 ± 0.0	0.2 ± 0.0
3,4-P-Xylose	1.6 ± 0.6	1.6 ± 0.2	1.3 ± 0.2	0.7 ± 0.1	1.0 ± 0.5	0.4 ± 0.1	0.8 ± 0.1	0.5 ± 0.2	0.4 ± 0.0	0.5 ± 0.2
2-X1-Xylose	1.5 ± 0.3	0.4 ± 0.2	0.3 ± 0.0	0.9 ± 0.3	0.5 ± 0.2	0.3 ± 0.1	0.4 ± 0.2	0.5 ± 0.1	0.3 ± 0.0	0.4 ± 0.0
T-F-Arabinose	21.3 ± 0.1	35.5 ± 2.2	27.6 ± 0.7	24.8 ± 1.5	13.6 ± 1.0	6.4 ± 0.5	9.7 ± 0.4	7.9 ± 0.4	12.3 ± 1	10.3 ± 0.9
5-F-Arabinose	1.8 ± 0.4	9.0 ± 1.7	6.1 ± 0.2	0.6 ± 0.2	0.9 ± 0.1	2.0 ± 0.6	1.0 ± 0.8	0.5 ± 0.2	0.6 ± 0.0	0.4 ± 0.2
3-Arabinose	0.8 ± 0.1	0.5 ± 0.1	0.8 ± 0.1	0.7 ± 0.1	1.1 ± 0.3	0.4 ± 0.1	0.8 ± 0.1	0.7 ± 0.1	1.2 ± 0.0	0.8 ± 0.0
2-F-Arabinose	0.9 ± 0.1	0.8 ± 0.1	1.4 ± 0.1	0.8 ± 0.1	0.8 ± 0.3	0.4 ± 0.1	0.8 ± 0.0	0.9 ± 0.1	1.2 ± 0.1	0.7 ± 0.0
2-X2-Arabinose	0.2 ± 0.1	0.2 ± 0.0	0.1 ± 0.0	0.2 ± 0.0	0.3 ± 0.1	0.1 ± 0.0	0.2 ± 0.0	0.2 ± 0.0	0.2 ± 0.1	0.2 ± 0.0
T-Fucose	1.0 ± 0.1	1.1 ± 0.1	1.4 ± 0.1	1.1 ± 0.0	0.9 ± 0.0	1.2 ± 0.0	0.9 ± 0.1	1.5 ± 0.1	1.4 ± 0.0	1.2 ± 0.1
T-Rhamnose	0.2 ± 0.0	0.5 ± 0.1	3.0 ± 0.2	0.2 ± 0.0	1.4 ± 0.1	1.4 ± 0.1	2.2 ± 0.1	1.5 ± 0.1	3.3 ± 0.2	2.0 ± 0.1
2-Rhamnose	0.0 ± 0.0	0.3 ± 0.1	0.5 ± 0.0	0.1 ± 0.0	0.4 ± 0.1	0.5 ± 0.2	0.7 ± 0.2	0.5 ± 0.2	1.5 ± 0.2	0.7 ± 0.3
T-GlcA	0.1 ± 0.0	0.1 ± 0.0	0.1 ± 0.0	0.1 ± 0.0	0.1 ± 0.0	0.1 ± 0.0	0.1 ± 0.0	0.1 ± 0.0	0.1 ± 0.0	0.1 ± 0.0
T-GalA	0.2 ± 0.0	0.2 ± 0.0	0.2 ± 0.0	0.2 ± 0.0	0.2 ± 0.0	0.1 ± 0.0	0.1 ± 0.0	0.1 ± 0.0	0.2 ± 0.0	0.2 ± 0.0
T-Mannose	2.8 ± 0.3	1.8 ± 0.1	2.3 ± 0.1	3.0 ± 0.2	1.9 ± 0.2	1.6 ± 0.3	1.5 ± 0.0	1.6 ± 0.2	3.3 ± 0.4	2.7 ± 0.2
4-Mannose/3-Mannose	4.8 ± 1.0	2.1 ± 0.3	3.2 ± 0.2	3.7 ± 0.6	10.9 ± 1.2	3.8 ± 0.2	7.3 ± 0.5	6.6 ± 0.5	4.3 ± 0.5	6.6 ± 0.4
2-Mannose	0.4 ± 0.1	0.3 ± 0.0	0.4 ± 0.0	0.4 ± 0.2	0.2 ± 0.0	0.1 ± 0.1	0.1 ± 0.1	0.2 ± 0.0	0.3 ± 0.0	0.3 ± 0.0
4,6-Mannose	0.3 ± 0.1	0.1 ± 0.0	0.3 ± 0.0	0.1 ± 0.0	0.5 ± 0.2	0.1 ± 0.0	0.2 ± 0.0	0.3 ± 0.0	0.2 ± 0.0	0.1 ± 0.0
3,6-Mannose	0.5 ± 0.1	0.3 ± 0.1	0.5 ± 0.0	0.6 ± 0.0	0.2 ± 0.0	0.2 ± 0.0	0.2 ± 0.0	0.2 ± 0.0	0.3 ± 0.0	0.2 ± 0.0
X-Hexose	1.3 ± 0.2	2.0 ± 0.2	1.8 ± 0.2	1.5 ± 0.2	1.2 ± 0.1	1.1 ± 0.2	1.4 ± 0.1	1.2 ± 0.0	1.9 ± 0.1	1.2 ± 0.1
X,X-Hexose (I)	0.9 ± 0.1	0.5 ± 0.2	0.5 ± 0.2	0.5 ± 0.2	1.0 ± 0.6	0.5 ± 0.2	0.8 ± 0.3	1.1 ± 0.4	0.4 ± 0.1	0.5 ± 0.2
2,X-Hexose (I)	0.4 ± 0.1	0.2 ± 0.1	0.1 ± 0.0	0.3 ± 0.1	0.7 ± 0.4	0.6 ± 0.1	0.6 ± 0.3	0.7 ± 0.3	0.4 ± 0.1	0.3 ± 0.1
2,X,X-Hexose (I)	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.1 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.1 ± 0.0	0.1 ± 0.0	0.0 ± 0.0
2,X,X-Hexose (II)	0.3 ± 0.2	0.1 ± 0.1	0.1 ± 0.0	0.2 ± 0.1	0.4 ± 0.2	0.2 ± 0.1	0.3 ± 0.1	0.3 ± 0.1	0.3 ± 0.0	0.2 ± 0.1
X-Deoxyhexose (I)	0.0 ± 0.0	0.1 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.1 ± 0.0	0.5 ± 0.1	0.5 ± 0.0	0.5 ± 0.0	0.4 ± 0.0	1.3 ± 0.2

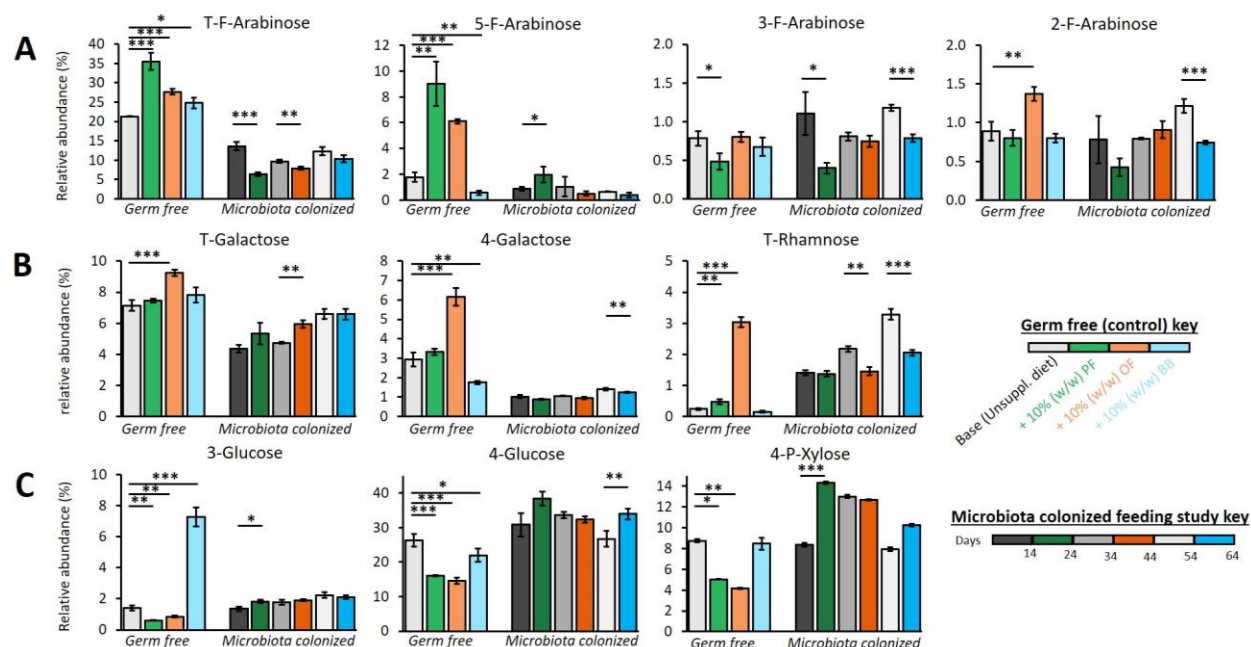|										
X-Deoxyhexose (II)	0.3 ± 0.1	0.3 ± 0.0	1.6 ± 0.1	0.2 ± 0.0	0.3 ± 0.0	0.4 ± 0.1	0.3 ± 0.0	1.3 ± 0.1	0.4 ± 0.0	0.4 ± 0.1
X-Deoxyhexose (III)	0.3 ± 0.1	0.2 ± 0.0	0.3 ± 0.1	0.1 ± 0.0	1.6 ± 0.5	2.0 ± 0.4	3.1 ± 0.2	2.4 ± 0.1	3.9 ± 0.5	2.8 ± 0.2

Figure 4.6. Monitoring the utilization of polysaccharides through relative linkage abundances in germ free and microbe-colonized feces. Linkages associated to (A) linear and branched arabinan, galactan from pectin (B), and β-glucan from bran (C). The results are represented by mean ± standard deviation of experimental replicates. The error bars indicate the standard deviation and P value corresponded to level of significance by t-test (*$P<0.05$, **$P<0.01$, ***$P<0.001$).

The linkages from fibers in GF and microbiota colonized mice feces were analyzed by comparing their relative abundances in bar graphs. **Figure 4.6A** displays the linkages associated with both linear and branched arabinan in the GF and microbiota colonized sample set. The T-arabinose in the pea and orange fibers were overly abundant in the GF control samples. However, in the 64-day feeding trial, both pea and orange fiber linkages significantly decreased.

Similarly, the 5-arabinose was found to be overly dominant in the GF samples and was utilized in the samples from the 64-day feeding trail. The 3-arabinose was not abundant in the GF samples, however 2-arabinose was for the orange fiber. The results indicate the arabinan structures in orange fiber had branched side chains $\alpha(1\rightarrow3)$ linkages whereas the pea fiber arabinan had minimal branching.

Glycosidic linkages from galactan were also monitored in the GF and microbe colonized samples. The T-galactose and 4-galactose showed statistically significant abundances in the GF samples (**Figure 4.6B**). The T-galactose linkages were less abundant in the feeding study; however, the linkages remained abundant compared to the unsupplemented base diet. The results suggested T-galactose from other sources or galactans are still present in the fecal sample. On the other hand, 4-galactose abundance was consumed in the microbe colonized samples. An abundant T-rhamnose was observed in the GF orange fiber sample and suggested the presence of RGI. The structure of RGI is known to contain galacturonic acid and rhamnose residues.[33] However, their analysis remains a challenge for reasons mentioned previously. Despite the GalA linkage limitations, rhamnose linkages corresponding to pectin polysaccharides were monitored.

An abundant 3-glucose coming from β-glucan was observed in the barley bran diet. In **Figure 4.6C**, 3-glucose was significantly represented in the GF feces and decreased in the microbe colonized samples. As mentioned from the text above, 3-glucose is typically present in brans. The 4-glucose was also present in β-glucan; however, the linkage did not have a significant difference between the GF and microbe colonized samples. This may have been caused by the other 4-glucose containing polysaccharides in the samples, such as starch and cellulose.

Linkage analysis in human feces

To obtain linkage profiles of human feces to monitor fiber-microbe interaction, the LC-MRM-MS method was applied to fecal samples from a human feeding study. A total of 18 human subjects participated in the feeding study. The study included a pre-intervention period during weeks 1- 2 without a fiber snack followed by weeks 2 - 5 and weeks 5 - 8 where pea fiber-snacks were administered three times a day. Furthermore, human participants were on their free base-line diets and not on HiSF-LoFV diet. For each participant, three fecal samples were collected at the end of week 1, week 5, and week 8. These collected fecal samples were subjected to the linkage analysis workflow.
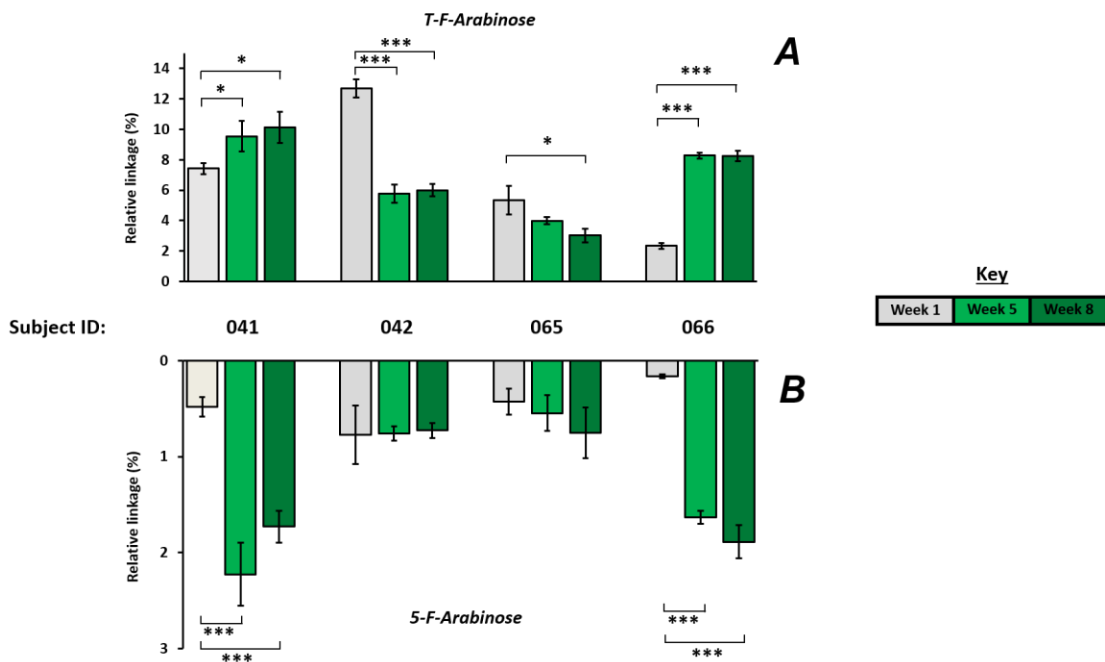

Figure 4.7. The relative linkage abundances of T-F-Arabinose (A) and 5-F-Arabinose (B) in the human feeding study of four human participants. Error bars corresponded to standard deviation

of experimental replicates (n=3). A two-tailed T-test was performed and P value corresponded to level of significance by t-test (*$P<0.05$, **$P<0.01$, ***$P<0.001$).

For control in the human study, the pre-intervention fecal sample was used as a evaluate whether the fiber-snack intervention was utilized by gut microbes for each subject. For example, the grey bars shown in **Figure 4.7**, represented the control for each participant and the green bars indicated the abundances of pea fiber linkages used to monitor fiber utilization in the feeding study. The linkage analysis of pea fiber revealed the presence of linear arabinan corresponding to T-arabinose and 5-arabinose. The utilization of arabinan in four subjects is demonstrated in **Figure 4.7**, where the abundances of arabinose linkages varied throughout the study. Human subject 041, yielded a statistically significant lower abundance for T-arabinose and 5-arabinose (7.41 % and 0.48 %) pre-intervention compared to the higher amounts of T-arabinose (9.53 % and 10.13 %) and 5-arabinose (2.23 % and 1.72 %) observed during fiber intervention in weeks 5 and 8. This indicated subject 041 did not utilized arabinan during the study and also provided insight that the gut microbes do not contain the GHs to depolymerize arabinan.

For subject 042, altered amounts of T-arabinose and 5-arabinose were observed throughout the feeding study. The abundances for T-arabinose decreased from 12.66 % for week 1 and to 5.76 % and 5.99 % for weeks 5 and week 8, respectively. However, we did not observe a significant difference in 5-arabinose from 0.77 % for week 1 to 0.76 % and 0.72% for weeks 5 and week 8, respectively. The results indicated gut microbes had preferential consumption for T-arabinose species while leaving behind the 5-arabinose substrates.

CONCLUSION

A liquid chromatography tandem mass spectrometry-based method was employed for the analysis of carbohydrate linkages in a large feeding study including fibers, diets, mice and human fecal biospecimens. The linkage analysis method was used to deduce the polysaccharide structures in fibers and were monitored throughout the feeding study. The method provided quantitation of carbohydrate linkages that allowed for the comparison of linkages between fecal samples distributed throughout the feeding study. The 96-well plate analysis greatly increased the throughput which allowed for processing many samples. The utility of the linkage analysis shows a promising approach to study food-microbe interactions in both *in vitro* and *in vivo* clinical studies.

REFERENCE

1.   Amicucci, M. J.; Nandita, E.; Lebrilla, C. B., Function without structures: the need for in-depth analysis of dietary carbohydrates. *Journal of Agricultural and Food chemistry* **2019,** *67* (16), 4418-4424.

2.   Hizukuri, S., Relationship between the distribution of the chain length of amylopectin and the crystalline structure of starch granules. *Carbohydrate Research* **1985,** *141* (2), 295-306.

3.   Wang, L.; White, P., Structure and properties of amylose, amylopectin. *Cereal Chem* **1994,** *71* (3), 263-268.

4.   Butterworth, P. J.; Warren, F. J.; Ellis, P. R., Human α-amylase and starch digestion: An interesting marriage. *Starch-Stärke* **2011,** *63* (7), 395-405.

5.   Jarvis, M. C.; McCann, M. C., Macromolecular biophysics of the plant cell wall: concepts and methodology. *Plant Physiology and Biochemistry* **2000,** *38* (1-2), 1-13.

6.   Yatsunenko, T.; Rey, F. E.; Manary, M. J.; Trehan, I.; Dominguez-Bello, M. G.; Contreras, M.; Magris, M.; Hidalgo, G.; Baldassano, R. N.; Anokhin, A. P., Human gut microbiome viewed across age and geography. *Nature* **2012,** *486* (7402), 222-227.

7.   Marchesi, J. R.; Adams, D. H.; Fava, F.; Hermes, G. D.; Hirschfield, G. M.; Hold, G.; Quraishi, M. N.; Kinross, J.; Smidt, H.; Tuohy, K. M., The gut microbiota and host health: a new clinical frontier. *Gut* **2016,** *65* (2), 330-339.

8.   Takahashi, K.; Nishida, A.; Fujimoto, T.; Fujii, M.; Shioya, M.; Imaeda, H.; Inatomi, O.; Bamba, S.; Andoh, A.; Sugimoto, M., Reduced abundance of butyrate-producing bacteria species in the fecal microbial community in Crohn's disease. *Digestion* **2016,** *93* (1), 59-65.

9.   Qin, J.; Li, Y.; Cai, Z.; Li, S.; Zhu, J.; Zhang, F.; Liang, S.; Zhang, W.; Guan, Y.; Shen, D., A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **2012,** *490* (7418), 55-60.

10.   Zmora, N.; Suez, J.; Elinav, E., You are what you eat: diet, health and the gut microbiota. *Nature Reviews Gastroenterology & Hepatology* **2019,** *16* (1), 35-56.

11.   David, L. A.; Maurice, C. F.; Carmody, R. N.; Gootenberg, D. B.; Button, J. E.; Wolfe, B. E.; Ling, A. V.; Devlin, A. S.; Varma, Y.; Fischbach, M. A., Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **2014,** *505* (7484), 559-563.

12.   Rahman, M. S.; Kang, I.; Lee, Y.; Habib, M. A.; Choi, B. J.; Kang, J. S.; Park, D.-S.; Kim, Y.-S., *Bifidobacterium longum* subsp. infantis YB0411 Inhibits Adipogenesis in 3T3-L1 Pre-adipocytes and Reduces High-Fat-Diet-Induced Obesity in Mice. *Journal of Agricultural and Food Chemistry* **2021**.

13.   Lombard, V.; Golaconda Ramulu, H.; Drula, E.; Coutinho, P. M.; Henrissat, B., The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic acids research* **2014,** *42* (D1), D490-D495.

14.   El Kaoutari, A.; Armougom, F.; Gordon, J. I.; Raoult, D.; Henrissat, B., The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nature Reviews Microbiology* **2013,** *11* (7), 497-504.

15.   Englyst, H.; Wiggins, H. S.; Cummings, J., Determination of the non-starch polysaccharides in plant foods by gas-liquid chromatography of constituent sugars as alditol acetates. *Analyst* **1982,** *107* (1272), 307-318.

16.   Doco, T.; O'neill, M.; Pellerin, P., Determination of the neutral and acidic glycosyl-residue compositions of plant polysaccharides by GC-EI-MS analysis of the trimethylsilyl methyl glycoside derivatives. *Carbohydrate Polymers* **2001,** *46* (3), 249-259.

17.   McGinnis, G. D.; Biermann, C. J., *Analysis of monosaccharides as per-O-acetylated aldononitrile (PAAN) derivatives by gas–liquid chromatography (GLC)*. CRC Press, Boca Raton, FL: 1989.

18.   Galermo, A. G.; Nandita, E.; Barboza, M.; Amicucci, M. J.; Vo, T.-T. T.; Lebrilla, C. B., Liquid chromatography–tandem mass spectrometry approach for determining glycosidic linkages. *Analytical Chemistry* **2018,** *90* (21), 13073-13080.

19.   Galermo, A. G.; Nandita, E.; Castillo, J. J.; Amicucci, M. J.; Lebrilla, C. B., Development of an extensive linkage library for characterization of carbohydrates. *Analytical Chemistry* **2019,** *91* (20), 13022-13031.

20.   Delannoy-Bruno, O.; Desai, C.; Raman, A. S.; Chen, R. Y.; Hibberd, M. C.; Cheng, J.; Han, N.; Castillo, J. J.; Couture, G.; Lebrilla, C. B., Evaluating microbiome-directed fibre snacks in gnotobiotic mice and humans. *Nature* **2021**, 1-5.

21.   Martín-Cabrejas, M. A.; Ariza, N.; Esteban, R.; Mollá, E.; Waldron, K.; López-Andréu, F. J., Effect of germination on the carbohydrate composition of the dietary fiber of peas (*Pisum sativum* L.). *Journal of Agricultural and Food Chemistry* **2003,** *51* (5), 1254-1259.

22.   Ratnayake, W. S.; Hoover, R.; Warkentin, T., Pea starch: composition, structure and properties—a review. *Starch-Stärke* **2002,** *54* (6), 217-234.

23.   Liwanag, A. J. M.; Ebert, B.; Verhertbruggen, Y.; Rennie, E. A.; Rautengarten, C.; Oikawa, A.; Andersen, M. C.; Clausen, M. H.; Scheller, H. V., Pectin biosynthesis: GALS1 in *Arabidopsis thaliana* is a β-1, 4-galactan β-1, 4-galactosyltransferase. *The Plant Cell* **2012,** *24* (12), 5024-5036.

24.   Maxwell, E. G.; Belshaw, N. J.; Waldron, K. W.; Morris, V. J., Pectin–an emerging new bioactive food polysaccharide. *Trends in Food Science & Technology* **2012,** *24* (2), 64-73.

25.   Daas, P. J.; Voragen, A. G.; Schols, H. A., Characterization of non-esterified galacturonic acid sequences in pectin with endopolygalacturonase. *Carbohydrate Research* **2000,** *326* (2), 120-129.

26.   Krall, S. M.; McFeeters, R. F., Pectin hydrolysis: effect of temperature, degree of methylation, pH, and calcium on hydrolysis rates. *Journal of Agricultural and Food Chemistry* **1998,** *46* (4), 1311-1315.

27.   Mohnen, D., Pectin structure and biosynthesis. *Current Opinion in Plant Biology* **2008,** *11* (3), 266-277.

28.   Morgan, K. R.; Ofman, D. J., Glucagel, a gelling β-glucan from barley. *Cereal Chemistry* **1998,** *75* (6), 879-881.

29.   Malunga, L. N.; Beta, T., Antioxidant capacity of water-extractable arabinoxylan from commercial barley, wheat, and wheat fractions. *Cereal Chemistry* **2015,** *92* (1), 29-36.

30.   Guo, R.; Xu, Z.; Wu, S.; Li, X.; Li, J.; Hu, H.; Wu, Y.; Ai, L., Molecular properties and structural characterization of an alkaline extractable arabinoxylan from hull-less barley bran. *Carbohydrate Polymers* **2019,** *218*, 250-260.

31.   Kabir, M.; Rizkalla, S. W.; Champ, M.; Luo, J.; Boillot, J.; Bruzzo, F. o.; Slama, G. r., Dietary amylose-amylopectin starch content affects glucose and lipid metabolism in adipocytes of normal and diabetic rats. *The Journal of Nutrition* **1998,** *128* (1), 35-42.

32.   Parks, B. W.; Nam, E.; Org, E.; Kostem, E.; Norheim, F.; Hui, S. T.; Pan, C.; Civelek, M.; Rau, C. D.; Bennett, B. J., Genetic control of obesity and gut microbiota composition in response to high-fat, high-sucrose diet in mice. *Cell Metabolism* **2013,** *17* (1), 141-152.

33.   Zdunek, A.; Pieczywek, P. M.; Cybulska, J., The primary, secondary, and structures of higher levels of pectin polysaccharides. *Comprehensive Reviews in Food Science and Food Safety* **2021,** *20* (1), 1101-1117.