Title

Development of Liquid Chromatography — Tandem Mass Spectrometry Methods for Food Glycomics

Author

Bacalzo, Nikita

Publication Date

2023

Peer reviewed|Thesis/dissertation

Development of Liquid Chromatography – Tandem Mass Spectrometry Methods For Food Glycomics

by

NIKITA P. BACALZO JR.

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

_____

William H. Casey

_____

Gang-yu Liu

Committee in Charge

2023

ABSTRACT

Carbohydrates are the most abundant biomolecule in nature and are involved mainly in the central metabolic processes of life and in providing structural integrity across the different domains of life. Carbohydrates also make up a major portion of our diets, with the low-molecular-weight and highly digestible carbohydrates serving as our main source of energy, while the high-molecular-weight and indigestible structures are carried to the distal gut where the gut microbes can utilize them. These gut microbes can in turn affect our physiology and health. Numerous studies have shown that dietary carbohydrates are implicated in some disease states, and that this can be mediated by the food-microbe-host interactions. It is therefore necessary that the analytical tools for carbohydrate characterization be able to speciate between the myriad structures of carbohydrates in food.

In this research work, a novel bottom-up LC-MS/MS-based glycomics method to characterize and quantify polysaccharides is presented. A non-enzymatic chemical reaction based on Fenton's chemistry was used to depolymerize polysaccharides into smaller oligosaccharide structures. In Chapter 1, an overview of carbohydrate structures and a survey of various techniques for carbohydrates analysis is provided. In Chapter 2, the polysaccharide analysis workflow was optimized and used to identify polysaccharides in various types of food, while Chapter 3 demonstrated how the workflow was improved to provide accurate and reproducible quantitation of multiple polysaccharides. In Chapter 4, an integrated multi-glycomics protocol was outlined in the most detailed way. This protocol contained three different LC-MS/MS-based glycomics analyses, namely monosaccharide-, linkage-, and polysaccharide

analyses. Several examples of applying these methods were also presented. Lastly, Chapter 5 provides the summary and the future implications of these analytical workflows in the field of carbohydrates research.

ACKNOWLEDGEMENTS

I would like to acknowledge and express my gratitude to Prof. Carlito B. Lebrilla for giving me the opportunity to work with different projects that have helped me grow as a person and as a scientist. I deeply appreciate his mentorship inside and outside of the lab, as well his patience and support to my scientific endeavors.

I would also like to thank our collaborators from various projects who have given us the chance to provide analytical support to cutting-edge studies: Jeffrey Gordon and team, David A. Mills, J. Bruce German, Danielle Lemay, Jennifer Smilowitz. I would like to thank them for sharing their expertise with me.

I like to thank all the members of Lebrilla League for their unending support and for making this academic journey memorable and worthwhile. Special thanks for the Nutrition group (Winnie, Matt, Eshani, Juan, Garret, Anita, Charlie, Chris, Shawn, Sophia) for being the best labmates anyone could ask for. I have learned so much from every one of you. I would also like to thank the undergraduate researchers who have helped me with the various projects.

I dedicate this work to my family (Nikita Sr., Victoria, Isiah, Nickolai, and Nichole), my bestfriends (Pierr, Louis, Paula, Anne, and Joey), my partner, Jessie, and his loving family (Esperanza, Raul, Priscilla, Edgar, and Raul Jr.), and our adorable pets (Minerva, Salem, Tweety, Sylvester, and Cherry).

Table of Contents

Chapter 1: Introduction to carbohydrates: structure, functions, analytical methods

Carbohydrates

The most abundant biomolecule in nature are carbohydrates. They are widely ubiquitous among all the domains of life, serving variety of biological functions, and comprised of different structural forms and variations. Monosaccharides are the simplest types of carbohydrates, and their chemical structure is the source of diversity in higher-order carbohydrates. Glucose is the most common monosaccharide in nature, and it plays a central role in cellular metabolism for any life form in the planet. The chemical functional groups in a monosaccharide are multiple hydroxyl groups and at least one carbonyl group. Aldoses are monosaccharides with aldehyde functional group, while ketoses have the ketone moiety.[1] Additionally, monosaccharides are also grouped according to the number of carbons they have, and each class consists of multiple stereoisomers. The aldopentose class for example has 3 stereocenters and so it has $2^3 = 8$ stereoisomers, while there are $2^4 = 16$ possible forms for aldohexoses. **Figure 1.1** shows the D configuration for the aldohexoses in Fischer projection. The L configuration corresponds to the mirror images of these compounds.

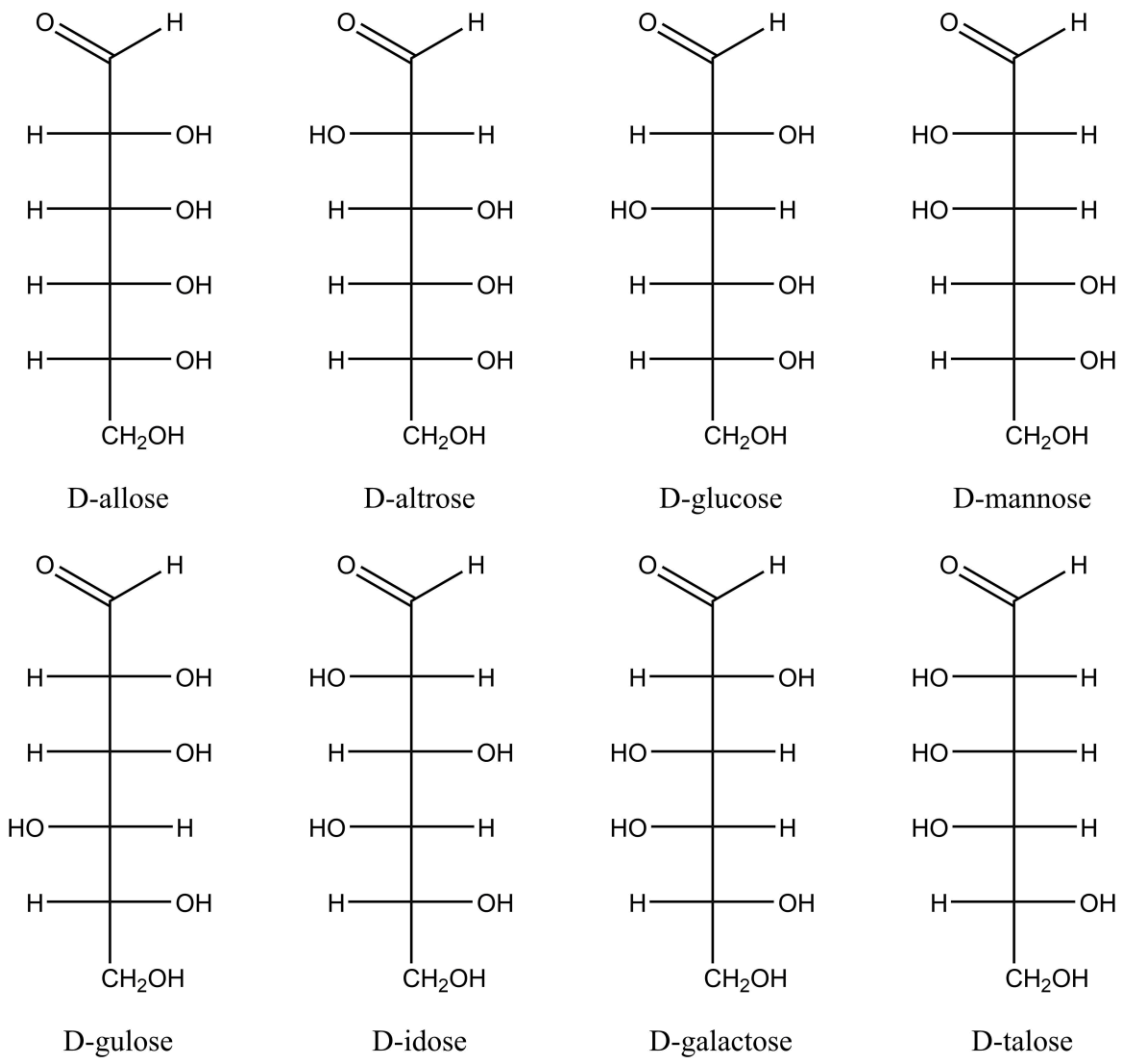


Figure 1.1. Fischer projections of the eight aldohexose monosaccharides in the D configuration.

Pentoses and hexoses usually exist in multiple closed-ring conformations because of the stability that 5- (furanose) or 6-carbon (pyranose) rings provide. Ring formation is facilitated by hemiacetal formation from a hydroxyl group and the carbonyl group. The resulting closed-ring structure can potentially have two anomeric configurations depending on the configuration of the hydroxyl group or hydrogen relative to the furthest stereogenic center (**Figure 1.2**). In solution, these structures, together with the acyclic form, are in equilibrium.[1,2]

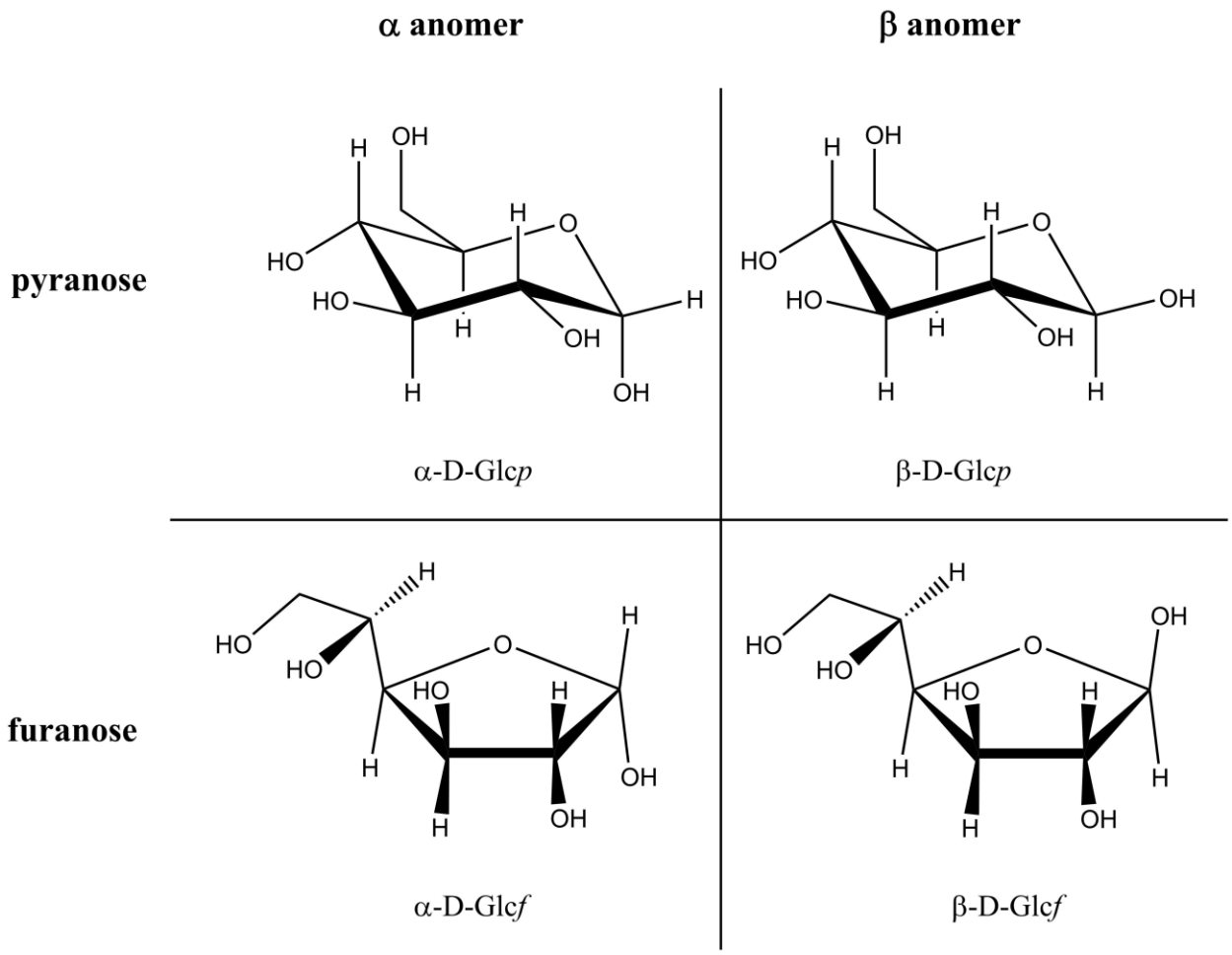


Figure 1.2. Different possible isomers of cyclic form of D-glucose.

With variations just on the level of monosaccharide alone, it is not surprising that carbohydrates have structural diversity that surpasses that of proteins and polynucleotides. However, the most commonly found monosaccharides in nature can be condensed into 14 monosaccharides as shown in **Figure 1.3**. These monosaccharides serve as the building blocks for higher-order structures such as oligo- and polysaccharides.[3]

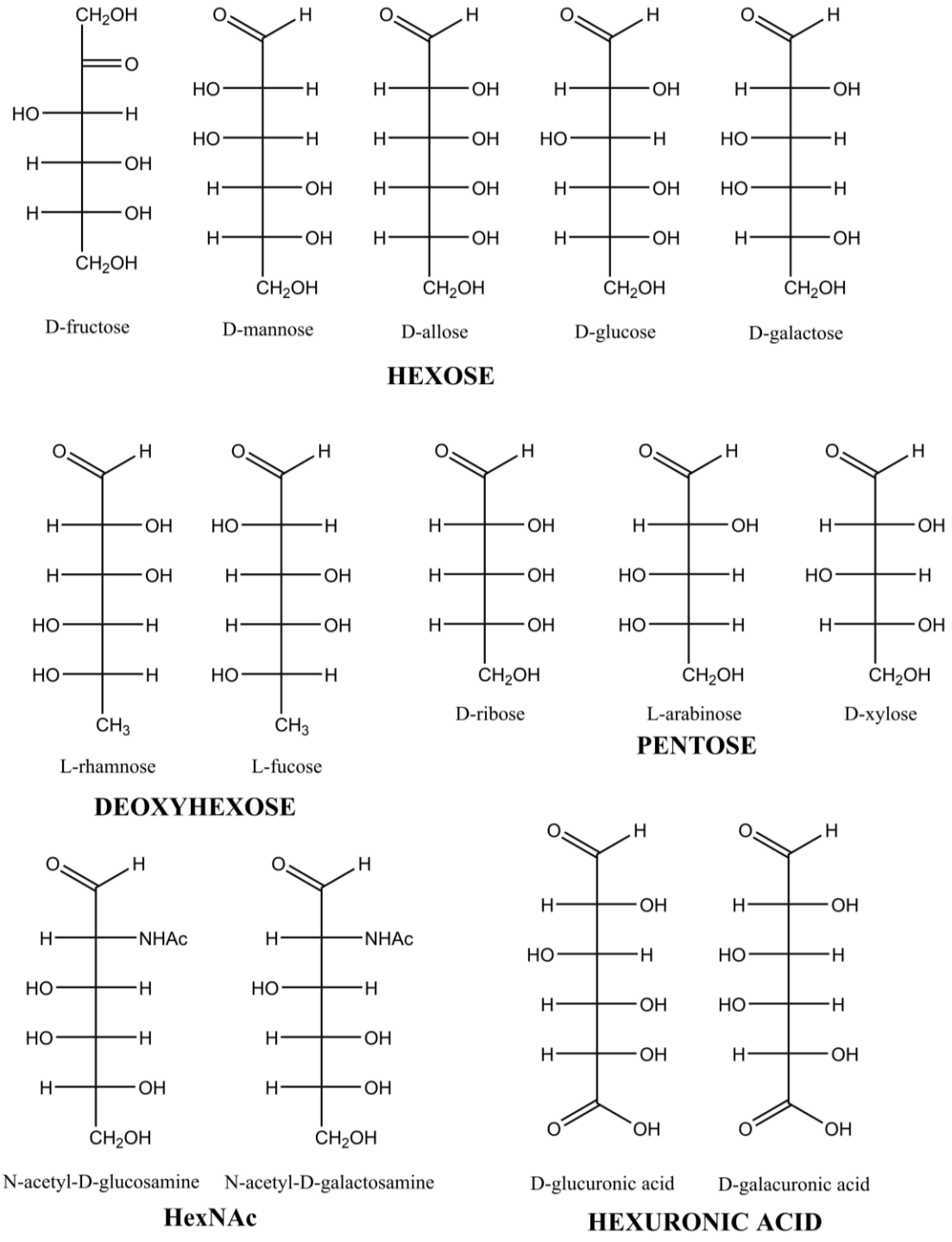


Figure 1.3. Fourteen most common monosaccharides found in nature.

Glycosidic linkages are formed between two monosaccharides, where the anomeric carbon from one monosaccharide is linked with a hydroxyl group of another monosaccharide by an oxygen atom. Glycosidic bond is an acetal formation reaction between the aldehyde group of a monosaccharide and a hydroxyl group of the other monosaccharide. Because there are two possible anomeric configurations, there are also two forms of glycosidic linkage.[2] Maltose and cellobiose are two examples where the only difference is the anomericity of the glycosidic linkage between the monosaccharides (**Figure 1.4**). These two disaccharides have very different physical and chemical properties.

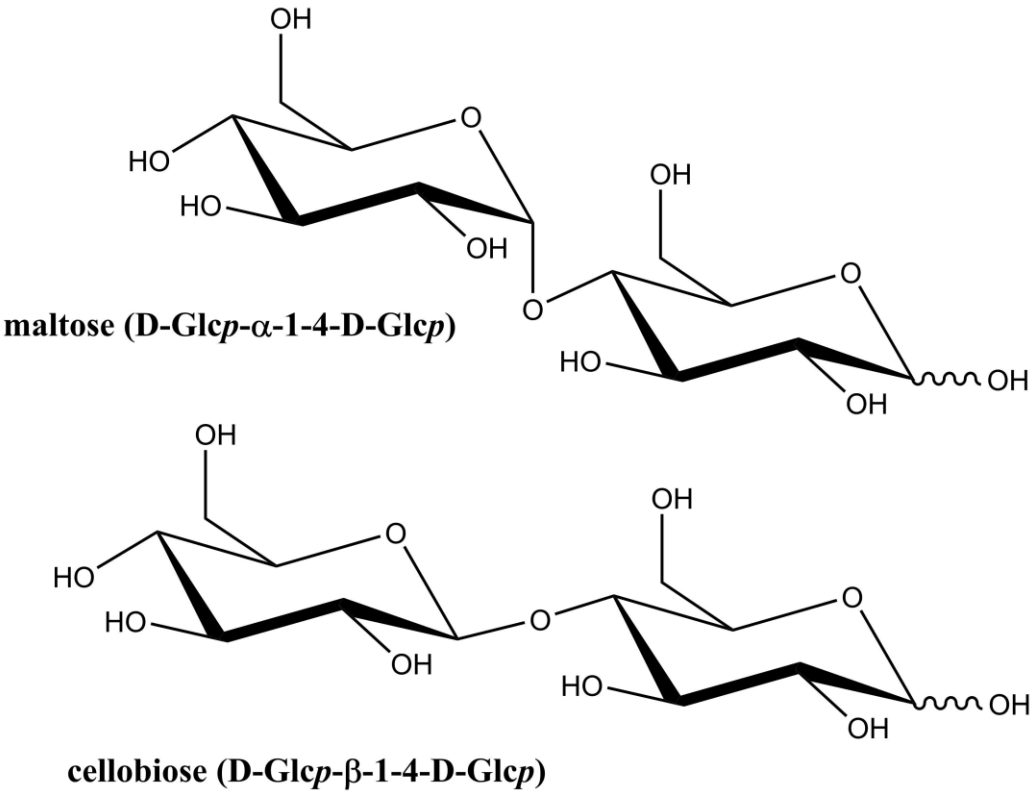


Figure 1.4. Two examples of glucose disaccharides with different anomeric configurations at the glycosidic linkage bond.

Additionally, multiple different hydroxyl groups can be linked to the anomeric carbon, resulting to multiple possible linkages. Two β-D-Glc*p* molecules can be linked via different linkages (**Figure 1.5**).

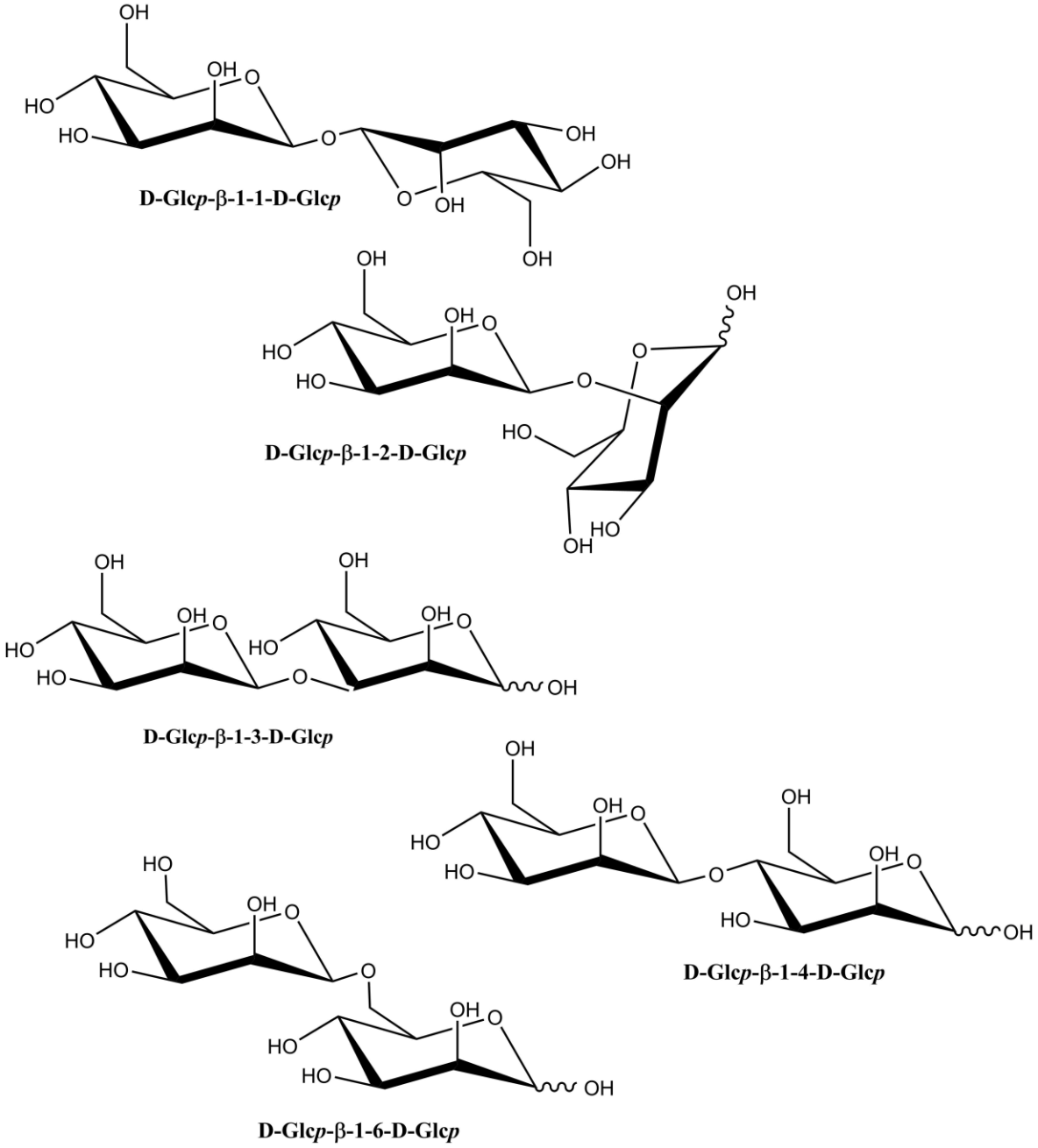


Figure 1.5. Multiple possible linkages between two D-Glc*p* monosaccharides.

Along with multitude possible monosaccharides, two anomeric configurations, and different multiple hydroxyl groups to form a linkage, the chemical diversity of all possible oligo- and polysaccharides structures is vast. For example, a disaccharide containing 2 aldohexose units can be linked via 10 linkage types, and each position can have 16 possible monosaccharides. With all these possibilities, there are 2,560 structures for an aldohexose disaccharide.

Plant Carbohydrates

Carbohydrates are highly abundant in plant, accounting for at least 50% of its dry weight.[4] As photosynthetic organisms, plants can synthesize carbohydrates from atmospheric CO_2 and water. It is then natural for these plants to use these different forms of carbohydrates in many of its physiological functions. Low-molecular-weight, highly soluble carbohydrates such as glucose, fructose, and sucrose, a disaccharide comprised of fructose and glucose, serve to distribute energy across different parts of the plant, specifically to non-photosynthetic parts.[5,6] High-molecular-weight polysaccharides serve as both energy storage and structural feature for plants. Starch, which is comprised of amylose and amylopectin, is the most abundant form of energy storage among different taxa of plants. Both amylose and amylopectin are homopolymers of α-D-Glc*p* with 1-4 linkages. Amylose is a linear chain with only α(1-4) linked D-Glc*p* while amylopectin is a branched structure with additional bisecting 4,6-linkages.[2,7] Fructans are another form of polysaccharide for energy storage. Fructans are made up of short chains of D-Fru*f*, usually with less than 100 DP. Inulin is a polymer with sucrose connected to a linear chain of 1-2 linked D-Fru*f* monosaccharides. Levan, on the other hand, has 2-6 linkage.[2,8,9] These energy storage polysaccharide structures are shown in **Figure 1.6**.

Another primary function of polysaccharides in plant is to provide structure and rigidity. This function is mainly carried out by cellulose, hemicelluloses, and pectins, which are all found in plant cell walls at different proportions among different tissue types and plant taxa. Cellulose is a linear polymer of β-D-Glc*p* with 1-4 linkage, similar to amylose but with different anomericity. Cellulose fibers are assembled as microfibrils, synthesized directly at the plasma membrane. The orientation of the hydroxyls allows the individual cellulose chains to arrange into crystalline structures. This crystallization makes cellulose one of the most chemically resistant polysaccharides.[10,11]

Hemicelluloses have similar structure with celluloses and contain β-1-4 linkages. For hemicelluloses, the linear backbone is usually comprised of D-Glc*p*, D-Man*p*, and D-Xyl*p*, which all have their C4 hydroxyl in the equatorial position.[2] Closely related to cellulose, xyloglucans also have the β-1-4-D-Glc*p* linear backbone with high degree of branching with α-1-6-D-Xyl*p*, which can also be capped with β-1-2-D-Gal*p* and α-1-2-L-Fuc*p*.[12,13] Mannans generally refer to any polysaccharide structures containing β-1-4 linked D-Man*p* as the linear backbone. Several polysaccharides fall under this structural motif such as linear mannans, galactomannans, and galactoglucomannans. Galactomannans have the linear mannan backbone with α-1-6-linked D-Gal*p* side groups. The ratio of mannose and galactose tend to vary across different types of plants. Glucomannan is a special mannan because it contains both D-Glc*p* and D-Man*p* in its linear backbone, which are then further branched with α-1-6-linked D-Gal*p*. Glucomannans also often have acetylation in its backbone.[2,14] Xylans refer to structures with linear backbone comprised of β-1-4-D-Xyl*p*. Glucuronoxylans have α-1-2-D-Glc*p*A substitution, which may also be methylated at O-4 position, and they are localized in the secondary cell walls of woody and herbaceous parts of the plant.[15,16] Arabinoxylans are found in the primary cell wall

and these structures have the linear xylan backbone with α-1-2- or α-1-3-L-Ara*f* substitutions. Additionally, like glucuronoxylan, arabinoxylan (a primary component of plant cell walls) may also have α-1-2-D-Glc*p*A and its 4-*O*-methylated counterpart. These arabinoxylans have more Ara*f* substitutions than their secondary cell wall counterpart.[2] These hemicelluloses are shown in **Figure 1.6**.

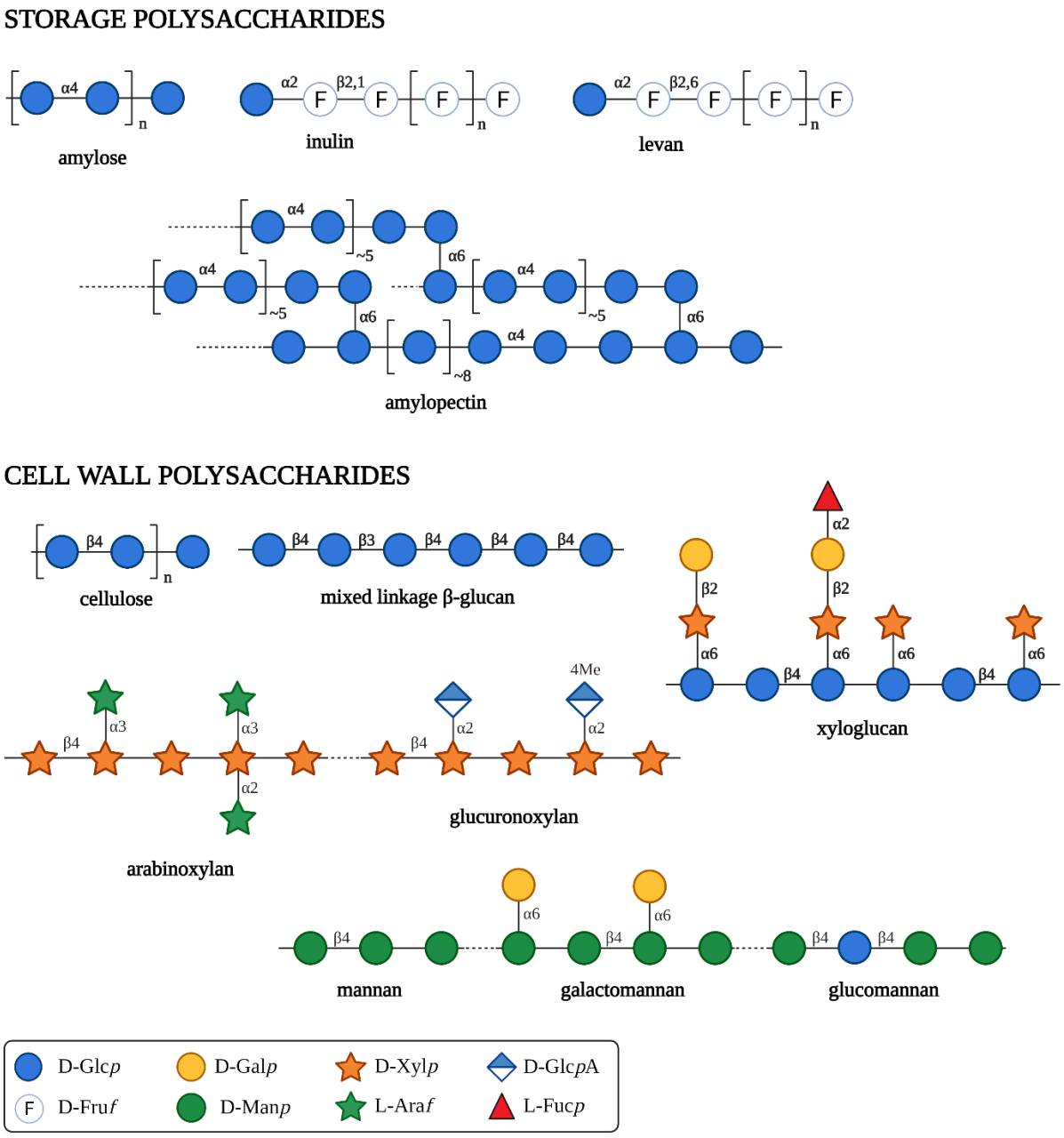


Figure 1.6. Examples of storage polysaccharides and hemicelluloses abundant in plants.

Pectins are one of the most structurally diverse and complicated polysaccharides in nature. These macromolecular complexes are found in the primary cell wall of plants and are comprised of various domains with some degree of conserved motifs. These domains are linked to each other both covalently and non-covalently.[2] Homogalacturonan is one of these domains comprised mainly of linear α-1-4-D-Gal*p*A backbone with varying degree of acetylation and methyl esterification.[17,18] Rhamnogalacturonan I (RG-I) is another domain with linear backbone comprised of repeating units α-1-2-L-Rha*p*-α-1-4-D-Gal*p*A. Additionally, the Rha*p* residues are further substituted with linear or branched chains of Ara*f* and Gal*p*, called pectic arabinans and pectic galactans, respectively.[2,19] Arabinans have α-1-5-L-Ara*f* linear backbone while galactans usually have β-1-4-D-Gal*p* linear backbone. These side chain domains are highly variable among different plant taxa. Pectic arabinans can be branched with α-1-3-L-Ara*f* substitution, which can further be extended with more α-1-5-L-Ara*f*. [20,21] Rhamnogalacturonan II (RG-II) is yet another pectin domain and is deemed to be the most structurally complicated polysaccharide comprised of 12 different monosaccharide residues and up to 21 glycosidic linkages. RG-II has a linear backbone comprised of α-1-4-D-Gal*p*Awith high degree of substitution ranging from mono-, di-, and oligosaccharide groups. Additionally, some of these residues are also O-methylated and O-acetylated. RG-II is almost exclusively in its dimer form, which is cross-linked by a bridging borate ester.[2,22] All these pectin domains, together with cellulose and hemicelluloses, contribute to the physical and chemical properties that it confers to the primary and secondary cell wall of plants.

Carbohydrates and the Gut Microbiota

Plants are the main source of carbohydrates in most human diets. Low-molecular-weight carbohydrates such as free glucose, fructose and sucrose can easily be digested and absorbed by the human gastrointestinal (GI) tract. These provide the necessary input for central metabolism. Other non-digestible carbohydrates pass through the GI tract intact until it reaches the colon where the gut microbiota is mostly localized. This non-digestable fraction is deemed as "dietary fiber", which is one of the macronutrients usually shown in food nutrition labels.[23]

The gut microbiota is a complex and dynamic ecosystem within the human GI tract comprised of up to 1000 species. These microbes are found in different proportions along the human GI tract starting from the mouth, totaling up to 10^{14} microorganisms. The colon has the highest density of anaerobic microbes.[24] Gut microbes play essential functions to the host health. They have been associated with several metabolic disorders, cardiovascular diseases, and some neurological symptoms.[25–28] However, in most of these associations, the underlying detailed mechanism of how the gut microbes contribute to the etiology of these diseases is still lacking. Many studies have indicated that link could be related to imbalances in the regulation of both local inflammation in the gut and the systemic inflammation.[28,29]

Oligo- and polysaccharides in the dietary fiber reach the colon where the gut microbiota can degrade them into mono-, di-, and smaller oligosaccharides. These degradation products are then used by the microbes to convert into other secondary metabolites such as short-chain fatty acids (SCFAs).[25] SCFAs have been shown to be critical factors in the homeostasis of the gut epithelial lining. Enterocytes can use them as energy source and signaling molecules. Some of these gut-derived metabolites can further enter circulation to affect other tissue systems of the body.[30,31]

The diversity of the gut microbiota stems from both the taxonomy as well as the functional enzymes that they express. It is estimated that the number of genes from the gut microbiota significantly exceeds that of the human genes by up to > 100 times.[24,25] Among these genes, carbohydrate-active enzymes (CAZymes) are the relevant players that the microbes use for the degradation and uptake of carbohydrates. CAZymes also include non-catalytic proteins such as carbohydrate-binding domains that aid in the recognition and binding of substrates.[32,33] Membrane-bound transporters are also necessary for the uptake of mono-, di-, and oligosaccharide across the cell wall and plasma membranes of the microbes. Inside the microbial cell, these low-molecular-weight carbohydrates can be further degraded into monosaccharides which are then essentially incorporated into several metabolic pathways.[34,35]

Given the critical importance of dietary fiber in the modulation of the gut microbiota, and that this specific interaction is mostly driven by the diversity of chemical structures of carbohydrates, it is therefore necessary to have analytical tools that can appropriately characterize food carbohydrates. However, conventional methods for carbohydrate analysis are limited in terms of coverage and rate of throughput. Next-generation sequencing workflows for metagenomics and metatranscriptomics has been used routinely for the detailed characterization of the gut microbial communities, enabling both taxonomic and functional profiling.[36–38] Additionally, given the high intra- and inter-individual variability between human beings, large cohort studies are often necessary to establish strong associations between nutritional interventions and gut microbial changes. High-throughput methodologies are then crucial to the success of these nutritional and clinical feeding trials.

Overview of analytical techniques for carbohydrates

Bulk measurements (enzymatic, gravimetric)

The most common way to measure bulk carbohydrate content specifically for food samples is mass by difference. Arguably, this may not be considered a measured value, rather a derived value from other attributes. The carbohydrates in most food nutrition labels are calculated by 100 minus the sum of protein, water, total fat, ash, and alcohol. This bulk value is generally taken as approximate, as the associated error is propagated from the other values used for its calculation. More importantly, this bulk value does not provide any kind of structural information of the carbohydrates in the food sample.

Dietary fibers can be measured using several AOAC methods, such as AOAC 991.43 and AOAC 2017.16.[39,40] In both methods, the first step involves *in vitro* digestion that simulates the condition of human digestion (**Figure 1.7**). This step includes treatment with amylase to breakdown starch and protease to break-down proteins. The resulting fraction is filtered, and the retentate is weighed and labeled as insoluble dietary fiber (IDF). The filtrate is then precipitated with ethanol and filtered again. The residue from the ethanol precipitation is the high-molecular-weight soluble dietary fiber (HMW SDF), while the filtrate is the low-molecular-weight soluble dietary fiber (LMW SDF). The main differences between the two AOAC methods are (1) the starch and protein hydrolysis conditions, and (2) the 991.43 method does not analyze the LMW SDF fraction. In AOAC 2017.16, the LMW SDF fraction is analyzed using any HPLC-based platforms. While these methods are validated and easy to implement, they are low-throughput and more importantly with no structural information provided. However, the fractions collected from these methods can be coupled with more informative analytical methods to fully characterize the carbohydrate structures from each different fraction.[41] It should also be noted

that the initial enzymatic digestion process is ideally close to the human digestion, however this assumption is likely unmet due to the complex nature of the human digestion process.

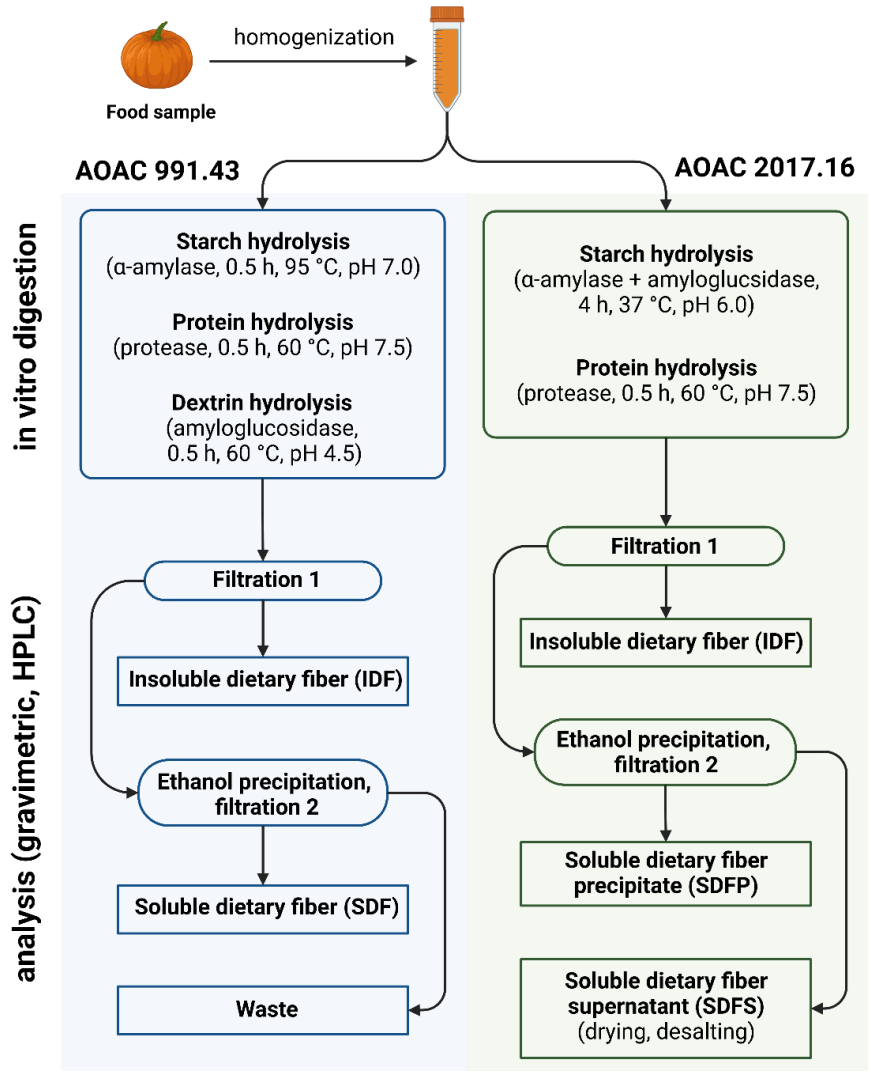


Figure 1.7. Analytical workflows for dietary fibers with AOAC 991.43 and 2017.16.

Nuclear magnetic resonance (NMR)

NMR remains the standard for complete structural elucidation of compounds including biomolecules such as carbohydrates. Small oligosaccharides can be easily analyzed with NMR as there are few overlapping peaks.[42] Additionally, multiple spectra from different nuclei and different pulse sequences are usually needed to corroborate and confirm peak assignments. Two-dimensional (2D) are also very helpful in providing more confident peak assignments. Provided that high quality spectra are obtained, highly detailed structural information can be generated from NMR analysis. This information includes absolute configuration (enantiomer configuration), monosaccharide identity (epimers), glycosidic linkage, substitutions and modifications (such as methylation, acetylation, phosphorylation), and anomericity.[43] However, acquiring high quality NMR spectra can be challenging, and several sample preparation steps are necessary to obtain a highly pure and concentrated sample. Because of this, NMR-based workflows tend to be low-throughput for generating carbohydrate composition profiles. Several studies have used NMR for quantitation of known oligosaccharides in various applications, such as CAZyme activity screening[44] and plant extract metabolomics[45] (including soluble mono-, di-, and oligosaccharides). For quantitation, an internal standard and multi-point calibration is usually necessary to get an accurate and precise absolute concentration of the analyte of interest.[46] One-point calibration may be possible for less complicated matrices.[47] Quantifier signal should also be confirmed to be non-overlapping with any other compounds in the matrix. Lastly, ^{1}H NMR analysis usually require high concentration samples (at least in the millimolar range) to generate high quality spectra, and ^{13}C NMR is even more less sensitive so analytes in the samples should at least be 5% w/w.[48] Several studies have attempted to use NMR to characterize and quantitate the more complex food polysaccharides.[49] Aside from the drawbacks

already mentioned, NMR signals from polysaccharides is more complicated and more difficult to annotate because the number of chemically independent nuclei exponentially increase with increasing DP.

Microarray-based platforms (lectins, monoclonal antibodies)

Lectins are proteins with have high binding affinity for certain carbohydrate structural motifs. Lectins are ubiquitous in nature serving various biological functions such as cell signaling, cell binding and adhesion, and immune regulation. Most of the characterized lectin ligands are motifs found in glycoprotein conjugates such as N- and O-glycans, and so these lectins are mostly used for probing glycoproteins and similar structures in various biological specimens such as tissue slides.[2,50] In a microarray format, samples are first deposited, usually on glass slides, which are then allowed to interact with different lectin probes. These probes are usually fluorescently labelled to enable sensitive visualization of the binding events.

Similarly, monoclonal antibodies (mAbs) which have been generated for specific carbohydrate epitopes have also been used as detection probes in microarray formats. The microarray format enables parallelization and high-throughput analysis, and fluorescence-based reporter provides a highly sensitive detection.[51] However, there are some limitations with microarray-based platforms. Samples usually need to be pre-processed first to obtain a high S/R signal. Carbohydrates can be extracted and fractionated by different solvents and buffers to separate based on their solubilities. In some cases, because of the structural similarity of carbohydrate molecules, these mAbs have some cross-reactivities. Therefore, a panel of multiple mAbs is usually used to confirm the structure of the unknown sample. Lastly, the lectin- and mAb-based microarrays are limited to relative quantitation because of the non-linearity of the readout response in using secondary probes.[52]

Monosaccharide and glycosidic linkage analysis

Both monosaccharide and linkage analyses rely on completely hydrolyzing the carbohydrates into its monomeric units and then analyzing the liberated monosaccharides. For monosaccharide analysis, samples are incubated with strong acid at elevated temperatures to ensure complete hydrolysis. The resulting monosaccharides can be then analyzed using various analytical platforms, such as gas chromatography coupled to flame ionization detector (GC-FID)[53] or mass spectrometry (GC-MS)[54], and liquid chromatography coupled to mass spectrometry (LC-MS)[3,55]. An advantage of the monosaccharide methods is its sensitivity for detection and quantitation. While GC-based platforms are conventionally used for monosaccharide analysis, the sample preparation involves more step and the requirement that analytes must be volatile limits the use of 96-well plates. The GC-based method requires an additional step of derivatization of the monosaccharides into their alditol acetate form. These derivatized monosaccharides are then analyzed using GC-MS with usual run time of 20 min.Single quadruple instruments commonly used in GC-MS tend to have lower S/R, and consequently lower sensitivity than LC-MS/MS-based methods. Recently, an improved method for monosaccharide analysis has been developed using LC-MS/MS. Sub-picomole detection has been reported using UPLC-QqQ MS with run time of < 5 min.[56,57] This monosaccharide analysis workflow includes acid hydrolysis with trifluoroacetic acid, labeling with 1-phenyl-3-methyl-5-pyrazolone (PMP,) chromatographic separation using 50-mm C18 stationary phase column, and detection with QqQ MS in dynamic multiple reaction monitoring mode (dMRM). Furthermore, this workflow was optimized for 96-well plate format, enabling high-throughput analysis of large batch of samples. Plants,[56] numerous food samples,[41,57] and biological samples (cecal and fecal samples)[37] have been analyzed for their monosaccharide composition using this workflow.

Because of the hydrolysis reaction, linkage and some structural information is lost in the monosaccharide analysis. Nonetheless, the accurate and precise quantitative information from this workflow is still valuable for characterization of carbohydrates.

Glycosidic linkage analysis is typically performed similar to the monosaccharide analysis with an additional step of permethylation. The permethylation step is usually carried out using the Hakomori method where free hydroxyl groups in the carbohydrate were first deprotonated by the methylsulfinyl carbanion, which was *in situ* generated by reaction of saturated NaOH solution and anhydrous dimethyl sulfoxide. Upon addition of iodomethane, the generated alkoxides are capped with methyl groups.[58] The subsequent steps follow the typical monosaccharide analysis workflow comprising of hydrolysis, labelling (optional), and instrumental analysis. Glycosidic linkages are determined by the pattern of permethylation of the liberated monosaccharides.[59] For example, linkage analysis of 2'-fucosyllactose, a common human milk oligosaccharide, is shown in **Figure 1.8**. Linkage analysis is conventionally done using GC-based instruments.[60] But recent methods developed using LC-MS/MS-based platforms have proven to be more sensitive in detection and more amenable for high-throughput analysis.[59,61] Linkage information provides more structural information than monosaccharide analysis alone. However, absolute quantitation with linkage analysis is yet to be developed. Combining both monosaccharide and linkage analyses provide comprehensive and detailed chemical information of carbohydrates.[41,56,62] Both monosaccharide and linkage analyses workflows using UPLC-QqQ platform are summarized in **Figure 1.8**.

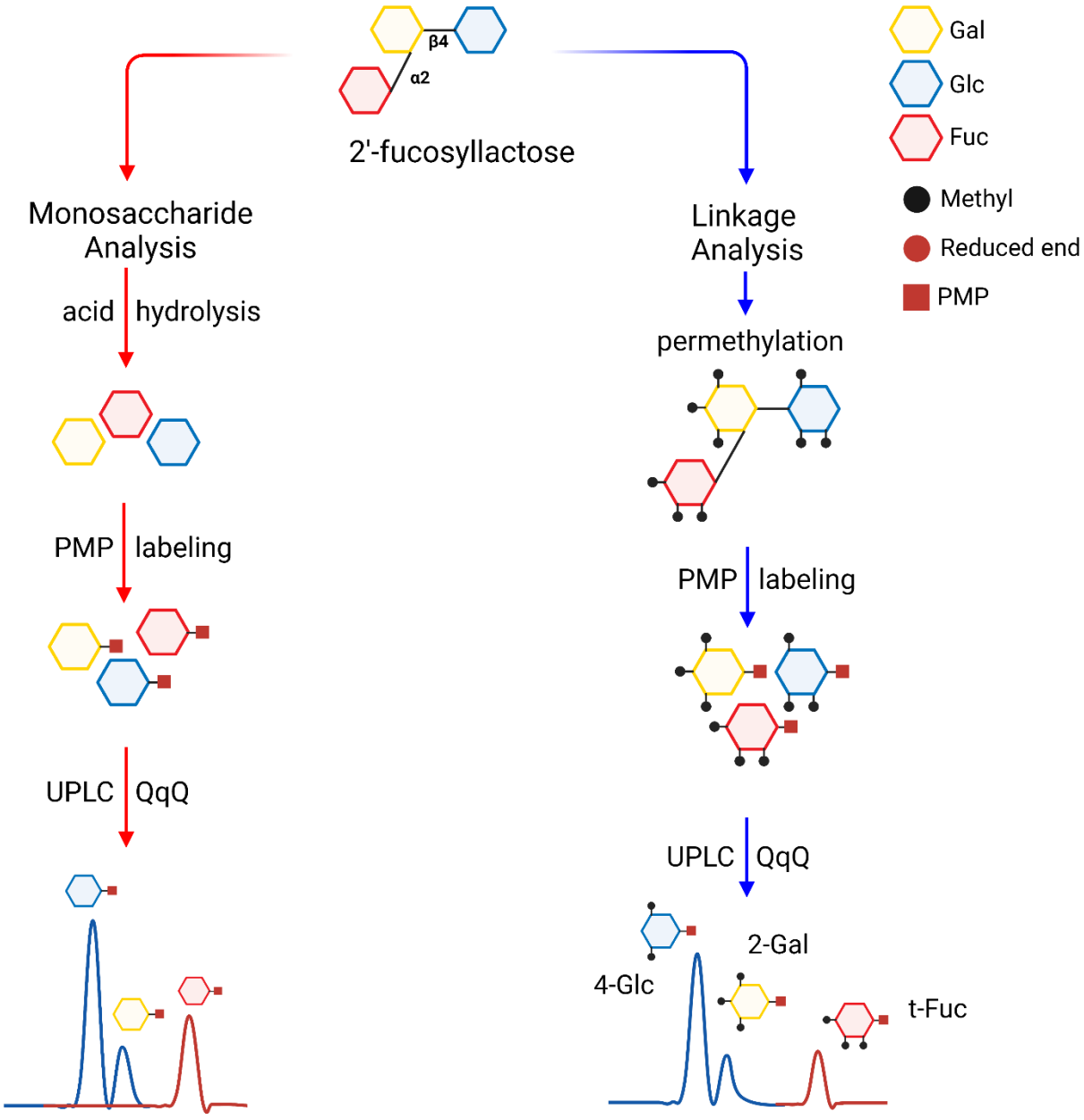


Figure 1.8. Schematic diagram for monosaccharide and glycosidic linkage analyses using UPLC-QqQ. 2'-fucosyllactose is used as an example.

Partial hydrolysis of polysaccharides for structural analysis

While monosaccharide and linkage analyses require the complete hydrolysis of carbohydrates into its monomeric units, partial hydrolysis using mild acid conditions[63] or enzyme digestion[64] produces smaller oligosaccharides. The resulting oligosaccharide fragments can then be analyzed using various instrumental approaches. This method is analogous to bottom-up proteomics and shotgun sequencing workflows, where enzymes are used to depolymerize proteins into peptides, and long nucleic acids into short oligonucleotides. However, there is no universal enzyme that can be used to depolymerize carbohydrates into shorter oligosaccharides. Controlled mild acid hydrolysis of carbohydrates generate short oligosaccharides fragments. Glycosyl hydrolase can also be used to cleave specific glycosidic linkages to provide structural and sequence information. However, prior information on the polysaccharide structure is required to know the enzymes suitable for generating oligosaccharides. In most cases, polysaccharides contain multiple glycosidic linkages and thus require several enzymes.[65] These different enzymes will need different optimum reaction conditions.

After generating the oligosaccharides, several platforms can be used to analyze the resulting oligosaccharide fragments. For a relatively pure sample with minimal matrix interference, matrix-assisted laser desorption ionization (MALDI)-based techniques has been used.[65] However, in most cases, the resulting mixture of oligosaccharides can be complicated and contain multiple isomers. Some common platforms include HPLC and capillary electrophoresis (CE). CE-based separation is based on multiple factors, primarily both the hydrodynamic volume and the charge state of the analytes. CE is usually coupled to absorbance- (UV or visible), fluorescence-, or MS-based detectors.[66] On the other hand, HPLC separates analytes primarily

based on its interaction with the stationary phase (analytical column) and the mobile phase. Similarly, HPLC can be coupled to various detection platforms as in CE. However, since oligosaccharides are generally non-fluorescent and non-chromoporic compounds, chemical labelling strategies are commonly used to be able to detect them using either fluorescence or absorbance.[67] MS-based detection requires analytes to be ionized in the gas phase. With the appropriate conditions, native oligosaccharides can be ionized with electrospray ionization (ESI). However, chemical labelling with a highly ionizable functional group significantly improves the detection with MS. Alternatively, native oligosaccharides have been analyzed using different forms of chromatography. High-performance anionic-exchange chromatography (HPAEC) uses high pH mobile phase and strong anion exchange resin to separate oligosaccharides, which requires significantly high pH (12-14) to be ionized. Pulsed amperometric detection (PAD) is conventionally used with HPAEC. Identification and quantification are done by using external calibration standards. However, isomeric separation can be challenging with HPAEC-PAD and so run times are usually longer than other HPLC methods.[68] Hydrophilic chromatography (HILIC) can also be used for native oligosaccharides where HILIC columns are usually based on amine, amide, or zwitterionic stationary phases. The mobile phase for HILIC is run in normal phase mode where organic solvent is the weaker eluent than high-water content solvents. HILIC can separate some isomeric structures but more complicated samples, such as containing partially hydrolyzed polysaccharide mixtures, can be challenging to analyze. Porous graphitic carbon (PGC) has been widely used for separation of oligosaccharides because of its superior isomeric separation.[69,70] In some cases, PGC can even separate anomeric structures. Both HILIC and PGC have been used for oligosaccharide profiling using MS-based detection.

Conclusions

Carbohydrates are structurally diverse biomolecules that contribute to a wide range of functionalities among the different domains of life. And true with most compounds known to man, their structure ultimately dictates their functions. And while carbohydrates are one of the major components of human diets, the analytical platforms to characterize food carbohydrates are still lacking and lagging behind other related methodologies. The overall aim of this dissertation is to provide novel workflows that can fill in this gap in analyzing food carbohydrates. Specifically, various HPLC-MS/MS based methods have been developed and used to gain a very detailed and comprehensive view of the carbohydrates in the foods that we eat.

Partial depolymerization is the main approach that was used in the polysaccharide analysis for the subsequent chapters in this dissertation. This non-enzymatic chemical reaction is called **F**enton's **I**nitiation **T**owards **D**efined **O**ligosaccharide **G**roups (FITDOG). Specifically, Fenton-based chemistry was used to generate reactive oxygen species, including hydroxyl radicals, which then cleaves the glycosidic linkages to produce smaller oligosaccharide fragments. HPLC-QqTOF MS was then used to profile the resulting oligosaccharide products. Using commercially available polysaccharide standards, a fingerprint library was generated. The FITDOG workflow was optimized and developed for polysaccharide identification in food samples (**Chapter 2**). Furthermore, the FITDOG workflow was improved and extended for simultaneous and absolute quantitation of polysaccharides in food samples (**Chapter 3**). Along with the monosaccharide and glycosidic linkage analyses, the FITDOG analysis was integrated in a multi-glycomics workflow to provide the most comprehensive and detailed carbohydrate characterization (**Chapter 4**). The multi-glycomics workflow include an ethanol precipitation

step during the sample preparation to separate low vs high molecular weight carbohydrates. The high-molecular-weight fraction containing the polysaccharides was subjected to monosaccharide, linkage, and FITDOG analyses.

References

(1) Berg, J. M.; Tymoczko, J. L.; Gatto, G. J.; Stryer, L. Carbohydrates. In *Biochemistry*; W.H. Freeman/McMillan Learning: New York, 2019.

(2) *Essentials of Glycobiology*, 4th editio.; Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Mohnen, D., Kinoshita, T., Packer, N. H., Prestegard, J. H., et al., Eds.; Cold Spring Harbor Laboratory Press, 2022.

(3) 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*, 200–207. https://doi.org/10.1039/c7an01530e.

(4) El Khadem, H. S. Carbohydrates. *Encycl. Phys. Sci. Technol.* **2003**, 369–416. https://doi.org/10.1016/B0-12-227410-5/00080-6.

(5) Stein, O.; Granot, D. An Overview of Sucrose Synthases in Plants. *Front. Plant Sci.* **2019**, *10* (February), 1–14. https://doi.org/10.3389/fpls.2019.00095.

(6) Yoon, J.; Cho, L. H.; Tun, W.; Jeon, J. S.; An, G. Sucrose Signaling in Higher Plants. *Plant Sci.* **2021**, *302*, 110703. https://doi.org/10.1016/J.PLANTSCI.2020.110703.

(7) Pfister, B.; Zeeman, S. C. Formation of Starch in Plant Cells. *Cell. Mol. Life Sci.* **2016**, *73* (14), 2781–2807. https://doi.org/10.1007/s00018-016-2250-x.

(8) Benkeblia, N. Insights on Fructans and Resistance of Plants to Drought Stress. *Front. Sustain. Food Syst.* **2022**, *6* (March), 1–12. https://doi.org/10.3389/fsufs.2022.827758.

(9) Nelson, C. J.; Spollen, W. G. Fructans. *Physiol. Plant.* **1987**, *71* (4), 512–516. https://doi.org/10.1111/j.1399-3054.1987.tb02892.x.

(10) Polko, J. K.; Kieber, J. J. The Regulation of Cellulose Biosynthesis in Plants. *Plant Cell* **2019**, *31* (2), 282–296. https://doi.org/10.1105/tpc.18.00760.

(11) Nishiyama, Y. Structure and Properties of the Cellulose Microfibril. *J. Wood Sci.* **2009**, *55* (4), 241–249. https://doi.org/10.1007/s10086-009-1029-1.

(12) Kim, S. J.; Chandrasekar, B.; Rea, A. C.; Danhof, L.; Zemelis-Durfee, S.; Thrower, N.; Shepard, Z. S.; Pauly, M.; Brandizzi, F.; Keegstra, K. The Synthesis of Xyloglucan, an Abundant Plant Cell Wall Polysaccharide, Requires CSLC Function. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117* (33), 20316–20324. https://doi.org/10.1073/PNAS.2007245117.

(13) Nishinari, K.; Takemasa, M.; Zhang, H.; Takahashi, R. Storage Plant Polysaccharides: Xyloglucans, Galactomannans, Glucomannans. In *Comprehensive glycoscience: from chemistry to systems biology*; Kamerling, H., Ed.; Elsevier, 2007; pp 613–652.

(14) Scheller, H. V.; Ulvskov, P. Hemicelluloses. *Annu. Rev. Plant Biol.* **2010**, *61*, 263–289. https://doi.org/10.1146/annurev-arplant-042809-112315.

(15) Wierzbicki, M. P.; Maloney, V.; Mizrachi, E.; Myburg, A. A. Xylan in the Middle: Understanding Xylan Biosynthesis and Its Metabolic Dependencies toward Improving

Wood Fiber for Industrial Processing. *Front. Plant Sci.* **2019**, *10* (February), 1–29. https://doi.org/10.3389/fpls.2019.00176.

(16) Pauly, M.; Gille, S.; Liu, L.; Mansoori, N.; de Souza, A.; Schultink, A.; Xiong, G. Hemicellulose Biosynthesis. *Planta* **2013**, *238* (4), 627–642. https://doi.org/10.1007/s00425-013-1921-1.

(17) Wolf, S.; Mouille, G.; Pelloux, J. Homogalacturonan Methyl-Esterification and Plant Development. *Mol. Plant* **2009**, *2* (5), 851–860. https://doi.org/10.1093/mp/ssp066.

(18) Du, J.; Anderson, C. T.; Xiao, C. Dynamics of Pectic Homogalacturonan in Cellular Morphogenesis and Adhesion, Wall Integrity Sensing and Plant Development. *Nat. Plants* **2022**, *8* (4), 332–340. https://doi.org/10.1038/s41477-022-01120-2.

(19) Kaczmarska, A.; Pieczywek, P. M.; Cybulska, J.; Zdunek, A. Structure and Functionality of Rhamnogalacturonan I in the Cell Wall and in Solution: A Review. *Carbohydr. Polym.* **2022**, *278*, 118909. https://doi.org/10.1016/J.CARBPOL.2021.118909.

(20) Yu, L.; Zhang, X.; Li, S.; Liu, X.; Sun, L.; Liu, H.; Iteku, J.; Zhou, Y.; Tai, G. Rhamnogalacturonan I Domains from Ginseng Pectin. *Carbohydr. Polym.* **2010**, *79* (4), 811–817. https://doi.org/10.1016/j.carbpol.2009.08.028.

(21) Cankar, K.; Kortstee, A.; Toonen, M. A. J.; Wolters-Arts, M.; Houbein, R.; Mariani, C.; Ulvskov, P.; Jorgensen, B.; Schols, H. A.; Visser, R. G. F.; et al. Pectic Arabinan Side Chains Are Essential for Pollen Cell Wall Integrity during Pollen Development. *Plant Biotechnol. J.* **2014**, *12* (4), 492–502. https://doi.org/10.1111/pbi.12156.

(22) O'Neill, M. A.; Ishii, T.; Albersheim, P.; Darvill, A. G. Rhamnogalacturonan II: Structure and Function of a Borate Cross-Linked Cell Wall Pectic Polysaccharide. *Annu. Rev. Plant Biol.* **2004**, *55*, 109–139. https://doi.org/10.1146/annurev.arplant.55.031903.141750.

(23) National Research Council (U.S.); Institute of Medicine (U.S.). Standing Committee on the Scientific Evaluation of Dietary Reference Intakes.; Institute of Medicine (U.S.). Panel on the Definition of Dietary Fiber. *Dietary Reference Intakes : Proposed Definition of Dietary Fiber*; National Academy Press: Washington, D.C., 2001.

(24) Thursby, E.; Juge, N. Introduction to the Human Gut Microbiota. *Biochem. J.* **2017**, *474* (11), 1823–1836. https://doi.org/10.1042/BCJ20160510.

(25) Valdes, A. M.; Walter, J.; Segal, E.; Spector, T. D. Role of the Gut Microbiota in Nutrition and Health. *BMJ* **2018**, *361*, 36–44. https://doi.org/10.1136/bmj.k2179.

(26) Turnbaugh, P. J.; Hamady, M.; Yatsunenko, T.; Cantarel, B. L.; Duncan, A.; Ley, R. E.; Sogin, M. L.; Jones, W. J.; Roe, B. A.; Affourtit, J. P.; et al. A Core Gut Microbiome in Obese and Lean Twins. *Nature* **2009**, *457* (7228), 480–484. https://doi.org/10.1038/nature07540.

(27) Tremlett, H.; Bauer, K. C.; Appel-Cresswell, S.; Finlay, B. B.; Waubant, E. The Gut Microbiome in Human Neurological Disease: A Review. *Ann. Neurol.* **2017**, *81* (3), 369–382. https://doi.org/10.1002/ana.24901.

(28) Witkowski, M.; Weeks, T. L.; Hazen, S. L. Gut Microbiota and Cardiovascular Disease.

Circ. Res. **2020**, *127* (4), 553–570. https://doi.org/10.1161/CIRCRESAHA.120.316242.

(29) Ma, W.; Nguyen, L. H.; Song, M.; Wang, D. D.; Franzosa, E. A.; Cao, Y.; Joshi, A.; Drew, D. A.; Mehta, R.; Ivey, K. L.; et al. Dietary Fiber Intake, the Gut Microbiome, and Chronic Systemic Inflammation in a Cohort of Adult Men. *Genome Med.* **2021**, *13* (1), 1–13. https://doi.org/10.1186/s13073-021-00921-y.

(30) Silva, Y. P.; Bernardi, A.; Frozza, R. L. The Role of Short-Chain Fatty Acids From Gut Microbiota in Gut-Brain Communication. *Front. Endocrinol. (Lausanne).* **2020**, *11* (January), 1–14. https://doi.org/10.3389/fendo.2020.00025.

(31) Deleu, S.; Machiels, K.; Raes, J.; Verbeke, K.; Vermeire, S. Short Chain Fatty Acids and Its Producing Organisms: An Overlooked Therapy for IBD? *EBioMedicine* **2021**, *66*. https://doi.org/10.1016/j.ebiom.2021.103293.

(32) Lombard, V.; Golaconda Ramulu, H.; Drula, E.; Coutinho, P. M.; Henrissat, B. The Carbohydrate-Active Enzymes Database (CAZy) in 2013. *Nucleic Acids Res.* **2014**, *42* (D1), 490–495. https://doi.org/10.1093/nar/gkt1178.

(33) Helbert, W.; Poulet, L.; Drouillard, S.; Mathieu, S.; Loiodice, M.; Couturier, M.; Lombard, V.; Terrapon, N.; Turchetto, J.; Vincentelli, R.; et al. Discovery of Novel Carbohydrate-Active Enzymes through the Rational Exploration of the Protein Sequences Space. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116* (13), 6063–6068. https://doi.org/10.1073/pnas.1815791116.

(34) Wardman, J. F.; Bains, R. K.; Rahfeld, P.; Withers, S. G. Carbohydrate-Active Enzymes (CAZymes) in the Gut Microbiome. *Nat. Rev. Microbiol.* **2022**, *20* (9), 542–556. https://doi.org/10.1038/s41579-022-00712-1.

(35) Pereira, G. V.; Abdel-Hamid, A. M.; Dutta, S.; D'Alessandro-Gabazza, C. N.; Wefers, D.; Farris, J. A.; Bajaj, S.; Wawrzak, Z.; Atomi, H.; Mackie, R. I.; et al. Degradation of Complex Arabinoxylans by Human Colonic Bacteroidetes. *Nat. Commun.* **2021**, *12* (1). https://doi.org/10.1038/s41467-020-20737-5.

(36) Peters, B. A.; Wilson, M.; Moran, U.; Pavlick, A.; Izsak, A.; Wechter, T.; Weber, J. S.; Osman, I.; Ahn, J. Relating the Gut Metagenome and Metatranscriptome to Immunotherapy Responses in Melanoma Patients. *Genome Med.* **2019**, *11* (1), 1–14. https://doi.org/10.1186/s13073-019-0672-4.

(37) 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.; et al. Evaluating Microbiome-Directed Fibre Snacks in Gnotobiotic Mice and Humans. *Nature* **2021**, *595* (7865), 91–95. https://doi.org/10.1038/s41586-021-03671-4.

(38) Abu-Ali, G. S.; Mehta, R. S.; Lloyd-Price, J.; Mallick, H.; Branck, T.; Ivey, K. L.; Drew, D. A.; Dulong, C.; Rimm, E.; Izard, J.; et al. Metatranscriptome of Human Faecal Microbial Communities in a Cohort of Adult Men. *Nat. Microbiol.* **2018**, *3* (3), 356–366. https://doi.org/10.1038/s41564-017-0084-4.

(39) McCleary, B. V.; DeVries, J. W.; Rader, J. I.; Cohen, G.; Prosky, L.; Mugford, D. C.; Champ, M.; Okuma, K. Determination of Insoluble, Soluble, and Total Dietary Fiber

(CODEX Definition) by Enzymatic-Gravimetric Method and Liquid Chromatography: Collaborative Study. *J. AOAC Int.* **2012**, *95* (3), 824–844. https://doi.org/10.5740/jaoacint.CS2011_25.

(40) McCleary, B. V. Total Dietary Fiber (Codex Definition) in Foods and Food Ingredients by a Rapid Enzymatic-Gravimetric Method and Liquid Chromatography: Collaborative Study, First Action 2017.16. *J. AOAC Int.* **2019**, *102* (1), 196–207. https://doi.org/10.5740/jaoacint.18-0180.

(41) Couture, G.; Luthria, D. L.; Chen, Y.; Bacalzo, N. P.; Tareq, F. S.; Harnly, J.; Phillips, K. M.; Pehrsson, P. R.; McKillop, K.; Fukagawa, N. K.; et al. Multi-Glycomic Characterization of Fiber from AOAC Methods Defines the Carbohydrate Structures. *J. Agric. Food Chem.* **2022**. https://doi.org/10.1021/acs.jafc.2c06191.

(42) Blundell, C. D.; Reed, M. A. C.; Overduin, M.; Almond, A. NMR Spectra of Oligosaccharides at Ultra-High Field (900 MHz) Have Better Resolution than Expected Due to Favourable Molecular Tumbling. *Carbohydr. Res.* **2006**, *341* (12), 1985–1991. https://doi.org/10.1016/j.carres.2006.05.017.

(43) Speciale, I.; Notaro, A.; Garcia-Vello, P.; Di Lorenzo, F.; Armiento, S.; Molinaro, A.; Marchetti, R.; Silipo, A.; De Castro, C. Liquid-State NMR Spectroscopy for Complex Carbohydrate Structural Analysis: A Hitchhiker's Guide. *Carbohydr. Polym.* **2022**, *277* (August 2021), 118885. https://doi.org/10.1016/j.carbpol.2021.118885.

(44) Irague, R.; Massou, S.; Moulis, C.; Saurel, O.; Milon, A.; Monsan, P.; Remaud-Siméon, M.; Portais, J. C.; Potocki-Véronèse, G. NMR-Based Structural Glycomics for High-Throughput Screening of Carbohydrate-Active Enzyme Specificity. *Anal. Chem.* **2011**, *83* (4), 1202–1206. https://doi.org/10.1021/ac1032148.

(45) Corol, D. I.; Ravel, C.; Rakszegi, M.; Gilles, C.; Bedo, Z.; Beale, M. H.; Shewry, P. R.; Ward, J. L. 1H-NMR Screening for the High-throughput Determination of Genotype and Environmental Effects on the Content of Asparagine in Wheat Grain. *Plant Biotechnol. J.* **2016**, *14*, 128–139. https://doi.org/doi: 10.1111/pbi.12364.

(46) Schievano, E.; Tonoli, M.; Rastrelli, F. NMR Quantification of Carbohydrates in Complex Mixtures. A Challenge on Honey. *Anal. Chem.* **2017**, *89* (24), 13405–13414. https://doi.org/10.1021/acs.analchem.7b03656.

(47) Stryeck, S.; Horvath, A.; Leber, B.; Stadlbauer, V.; Madl, T. NMR Spectroscopy Enables Simultaneous Quantification of Carbohydrates for Diagnosis of Intestinal and Gastric Permeability. *Sci. Rep.* **2018**, *8* (1), 1–8. https://doi.org/10.1038/s41598-018-33104-8.

(48) Giraudeau, P. Challenges and Perspectives in Quantitative NMR. *Magn. Reson. Chem.* **2017**, *55* (1), 61–69. https://doi.org/10.1002/mrc.4475.

(49) Merkx, D. W. H.; Westphal, Y.; van Velzen, E. J. J.; Thakoer, K. V.; de Roo, N.; van Duynhoven, J. P. M. Quantification of Food Polysaccharide Mixtures by 1H NMR. *Carbohydr. Polym.* **2018**, *179* (June 2017), 379–385. https://doi.org/10.1016/j.carbpol.2017.09.074.

(50) Yasuda, E.; Tateno, H.; Hirabarashi, J.; Iino, T.; Sako, T. Lectin Microarray Reveals

Binding Profiles of Lactobacillus Casei Strains in a Comprehensive Analysis of Bacterial Cell Wall Polysaccharides. *Appl. Environ. Microbiol.* **2011**, *77* (13), 4539–4546. https://doi.org/10.1128/AEM.00240-11.

(51) Wood, I. P.; Pearson, B. M.; Garcia-Gutierrez, E.; Havlickova, L.; He, Z.; Harper, A. L.; Bancroft, I.; Waldron, K. W. Carbohydrate Microarrays and Their Use for the Identification of Molecular Markers for Plant Cell Wall Composition. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (26), 6860–6865. https://doi.org/10.1073/pnas.1619033114.

(52) Rydahl, M. G.; Hansen, A. R.; Kračun, S. K.; Mravec, J. Report on the Current Inventory of the Toolbox for Plant Cell Wall Analysis: Proteinaceous and Small Molecular Probes. *Front. Plant Sci.* **2018**, *9* (May), 1–20. https://doi.org/10.3389/fpls.2018.00581.

(53) Ma, J.; Adler, L.; Srzednicki, G.; Arcot, J. Quantitative Determination of Non-Starch Polysaccharides in Foods Using Gas Chromatography with Flame Ionization Detection. *Food Chem.* **2017**, *220*, 100–107. https://doi.org/10.1016/J.FOODCHEM.2016.09.206.

(54) Xia, Y. G.; Sun, H. M.; Wang, T. L.; Liang, J.; Yang, B. Y.; Kuang, H. X. A Modified GC-MS Analytical Procedure for Separation and Detection of Multiple Classes of Carbohydrates. *Molecules* **2018**, *23* (6). https://doi.org/10.3390/molecules23061284.

(55) 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. https://doi.org/10.1016/j.ijms.2018.12.009.

(56) Couture, G.; Vo, T.-T. T.; Castillo, J. J.; Mills, D. A.; German, J. B.; Maverakis, E.; Lebrilla, C. B. Glycomic Mapping of the Maize Plant Points to Greater Utilization of the Entire Plant. *ACS Food Sci. Technol.* **2021**, *1* (11), 2117–2126. https://doi.org/10.1021/acsfoodscitech.1c00318.

(57) Castillo, J. J.; Couture, G.; Bacalzo, N. P.; Chen, Y.; Chin, E. L.; Blecksmith, S. E.; Bouzid, Y. Y.; Vainberg, Y.; Masarweh, C.; Zhou, Q.; et al. The Development of the Davis Food Glycopedia—A Glycan Encyclopedia of Food. *Nutrients* **2022**, *14* (8). https://doi.org/10.3390/nu14081639.

(58) Hakomori, S. I. A Rapid Permethylation of Glycolipid, and Polysaccharide Catalyzed by Methylsulfinyl Carbanion in Dimethyl Sulfoxide. *J. Biochem.* **1964**, *55* (2), 205–208. https://doi.org/10.1093/oxfordjournals.jbchem.a127869.

(59) 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. https://doi.org/10.1021/acs.analchem.9b03101.

(60) Lim, S. J.; Wan Aida, W. M.; Schiehser, S.; Rosenau, T.; Böhmdorfer, S. Structural Elucidation of Fucoidan from Cladosiphon Okamuranus (Okinawa Mozuku). *Food Chem.* **2019**, *272*, 222–226. https://doi.org/10.1016/J.FOODCHEM.2018.08.034.

(61) 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.

https://doi.org/10.1021/acs.analchem.8b04124.

(62) 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. *J. Am. Soc. Mass Spectrom.* **2021**, *32* (8), 2175–2185. https://doi.org/10.1021/jasms.1c00133.

(63) Xia, Y. G.; Yu, S. M.; Liang, J.; Yang, B. Y.; Kuang, H. X. Chemical Fingerprinting Techniques for the Differentiation of Polysaccharides from Genus Astragalus. *J. Pharm. Biomed. Anal.* **2020**, *178*, 112898. https://doi.org/10.1016/J.JPBA.2019.112898.

(64) Guan, J.; Li, S. P. Discrimination of Polysaccharides from Traditional Chinese Medicines Using Saccharide Mapping-Enzymatic Digestion Followed by Chromatographic Analysis. *J. Pharm. Biomed. Anal.* **2010**, *51* (3), 590–598. https://doi.org/10.1016/j.jpba.2009.09.026.

(65) Günl, M.; Gille, S.; Pauly, M. OLIgo Mass Profiling (OLIMP) of Extracellular Polysaccharides. *J. Vis. Exp.* **2010**, No. 40, 1–5. https://doi.org/10.3791/2046.

(66) Sun, X.; Lin, L.; Liu, X.; Zhang, F.; Chi, L.; Xia, Q.; Linhardt, R. J. Capillary Electrophoresis-Mass Spectrometry for the Analysis of Heparin Oligosaccharides and Low Molecular Weight Heparin. *Anal. Chem.* **2016**, *88* (3), 1937–1943. https://doi.org/10.1021/acs.analchem.5b04405.

(67) Ruhaak, L. R.; Zauner, G.; Huhn, C.; Bruggink, C.; Deelder, A. M.; Wuhrer, M. Glycan Labeling Strategies and Their Use in Identification and Quantification. **2010**, 3457–3481. https://doi.org/10.1007/s00216-010-3532-z.

(68) Mechelke, M.; Herlet, J.; Benz, J. P.; Schwarz, W. H.; Zverlov, V. V.; Liebl, W.; Kornberger, P. HPAEC-PAD for Oligosaccharide Analysis—Novel Insights into Analyte Sensitivity and Response Stability. *Anal. Bioanal. Chem.* **2017**, *409* (30), 7169–7181. https://doi.org/10.1007/s00216-017-0678-y.

(69) 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.

(70) Li, Q.; Xie, Y.; Wong, M.; Barboza, M.; Lebrilla, C. B. Comprehensive Structural Glycomic Characterization of the Glycocalyxes of Cells and Tissues. *Nat. Protoc.* **2020**, *15* (8), 2668–2704. https://doi.org/10.1038/s41596-020-0350-4.

Chapter 2. Polysaccharide Identification Through Oligosaccharide Fingerprinting

ABSTRACT

The identification of polysaccharide structures in complex samples remains a unique challenge complicated by the lack of specific tools for polymeric mixtures. In this work, we present a method that depolymerizes polysaccharides to generate diagnostic oligosaccharide markers that are then analyzed by high-performance liquid chromatography-quadrupole time-of-flight mass spectrometry (HPLC-QTOF MS). Rapid identification of food polysaccharides was performed by aligning the identified oligosaccharides with a library of oligosaccharide markers generated from standard polysaccharides. Measurements of standard and food polysaccharides were performed to obtain the contributions of the identified polysaccharides using percent peak coverage and angle cosine methods. The method was validated using a synthetic mixture of standard polysaccharides while the reproducibility was confirmed with experimental triplicates of butternut squash samples, where standard deviation was less than 3% for the relative abundance of oligosaccharides. The method was further employed to examine diverse set of food samples.

INTRODUCTION

Plant polysaccharides are the most abundant biomacromolecules found in nature, which serves as important structural components in the integrity of plant tissues (cellulose, hemicelluloses, pectins), as well as energy storage in the form of starch and fructans. Aside from their intrinsic biological functions, they are also central to a wide range of applications in nutrition and agriculture. Polysaccharides often have bioactive properties when consumed. For example, mushroom polysaccharides, which include chitin, α- and β-glucans, mannans, xylans, and galactans, are found to have antitumor and immunomodulating activities [1]. Beyond food, polysaccharides are also components in therapeutics and nutraceutical products [1,2]. Polysaccharides can also be used in monitoring agricultural products. Different stages in plant maturity have been associated with changes in polysaccharide compositions [3–5]. Apples, for example, are considered ripe when they reach a low glucan and high polyuronide polysaccharide composition [6]. Commercialization of polysaccharide products also employ polysaccharide composition analysis primarily for batch-to-batch product validation [7,8]. For example, the polysaccharide components in tea-based Chinese herbal medicines are monitored across batches to ensure all products contain the same composition in order to facilitate similar health benefits [9]. In such practices, the methods for polysaccharide analysis often requires tedious sample preparation and several instrument platforms, rendering them unsuitable for broad characterization of the different structural classes of polysaccharides.

There remains a clear and considerable need for rapid polysaccharide identification [10]. Plant polysaccharides have wide structural diversity as they can be composed of a variety of monosaccharides, linkage types, and degree of branching. As a result, there is no single method that can fully characterize polysaccharide compositions in complex matrices such as food. A

traditional method for structural elucidation of plant polysaccharides involves nuclear magnetic resonance spectroscopy [11]. However, this technique can only be performed on a highly concentrated and pure polysaccharide samples [12,13]. Plant polysaccharide composition analysis has also been performed by Fourier-transform infrared spectroscopy and applying extensive chemometrics [14,15]. In this analysis, most of the diagnostic peaks are in the fingerprint region of the spectrum, specifically in the 500-1500 cm^{-1} range. Even with the use of chemometrics, significant overlap in the fingerprint region causes the analysis to increase in complexity as the number of polysaccharides in the mixture increases.

Mass spectrometry (MS) has been employed for the characterization of plant polysaccharides due to its high sensitivity and the ability to discriminate by mass-to-charge ratios. However, to be amenable to MS, polysaccharides must be depolymerized into oligosaccharides through either enzymatic or chemical processes. For example, enzymatically released plant oligosaccharides were used as diagnostic fingerprints to identify polysaccharides. This method used matrix-assisted laser desorption ionization-time of flight MS to characterize the oligosaccharides [16,17]. However, the method lacks isomer separation of oligosaccharides which renders branching and linear structures indistinguishable. Additionally, a major drawback for enzymatic hydrolysis is that it requires the use of specific enzymes for each type of polysaccharide present. There is no single enzyme capable of universally cleaving all polysaccharides. Acid hydrolysis with gas-chromatography MS (GC-MS) of monosaccharides has also been used to predict the parent polysaccharide structures [18–21]. However, the monosaccharide arrangement information is lost during acid hydrolysis which renders the technique less suitable for overall polysaccharide identification. For example, both monosaccharide and linkage composition analyses cannot distinguish between amylose and cellulose, as both are (1→4)-linked glucose

polymers that differs only in the anomeric character of the linkage. Hence, a complementary method for intact polysaccharide analysis is still needed.

In this research, we employed a recently developed chemical method for the degradation of plant polysaccharides into oligosaccharides [22]. The method was optimized to be universal among plant polysaccharides. The resulting oligosaccharides were analyzed using liquid chromatography-mass spectrometry and were matched against a library of oligosaccharide fingerprints created from standard polysaccharides to determine the polysaccharide composition. Furthermore, the method was validated with commercially available polysaccharide standards and various food samples.

MATERIALS AND METHODS

Samples and materials

Sodium borohydride ($NaBH_4$), sodium hydroxide (NaOH), sodium acetate, glacial acetic acid, trifluoroacetic acid (TFA), hydrogen peroxide (H_2O_2), and iron(III) sulfate pentahydrate ($Fe_2(SO_4)_3 \cdot 5H_2O$) were purchased from Sigma-Aldrich (St. Louis, MO). Galactan, amylose, β-glucan, arabinan, xyloglucan, curdlan, arabinoxylan, lichenan, glucomannan, mannan, galactomannan, arabinogalactan and xylan were purchased from Megazyme (Bray, Ireland). Microcrystalline cellulose was purchased from ACROS Organics. Acetonitrile (ACN, HPLC grade) was purchased from Honeywell (Muskegon, MI). Formic Acid (FA) was purchased from Fisher Scientific (Belgium, UK). Porous graphitized carbon (PGC) solid phase extraction (SPE) plates were purchased from Glygen (Columbia, MD). Nanopure water (18.2 MΩ-cm) was used for all experiments. Yellow corn meal (*Zea mays*), wheat grass (*Triticum sp*.), whole grain oat

cereal (*Avena sativa*), horseradish root (*Armoracia rusticana*), and coffee grounds (*Coffea arabica*) were purchased from the Davis Co-op (Davis, CA). Coconut (*Cocos nucifera*), jackfruit (*Artocarpus heterophyllus*), guava (*Psidium guajava*), yam leaves (*Dioscorea sp*.), bok choy leaves (*Brassica rapa*) were purchased from 99 Ranch Market (Sacramento, CA). Coffee grounds were brewed using conventional hot water extraction and the spent coffee grounds were used for polysaccharide analysis.

Food and polysaccharide sample preparation

All food samples were freeze-dried and ground to fine powder using Bead Ruptor Elite Bead Mill Homogenizer from Omni International (Kennesaw, GA). Stock solutions of polysaccharide standards and food were prepared at 10 mg/mL aqueous suspension and homogenized further with the bead homogenizer. Samples were heated in an oven at 100 ^{o}C for 1 hr, and then 0.1 mL was transferred to reaction tubes. Aliquoted samples were dried by vacuum centrifugation prior to reaction.

Generation of representative oligosaccharides

Polysaccharides were depolymerized using an oxidative method that had been optimized for several polysaccharides [22]. The treatment was performed on standard polysaccharides and food samples. Briefly, a reaction solution was prepared by mixing 95 mL of 40 mM sodium acetate buffer adjusted to pH 5 with glacial acetic acid, 5 mL of 30% (v/v) hydrogen peroxide, and 3.2 mg $Fe_2(SO_4)_3{\cdot}(H_2O)_5$. The reaction mixture was vortexed and added to each dried polysaccharide standards and food samples resulting in a final concentration of 1 mg/mL. To initiate the reaction, samples were incubated at 100 °C for 20 mins with a follow-up treatment with half of the reaction

volume of 2 M NaOH to quench the reaction. Neutralization was performed by adding 61 µL of glacial acetic acid.

Reduction of oligosaccharides

Oligosaccharides were reduced by treatment with 1 M $NaBH_4$ for 1 hour of incubation at 65 ºC. Purification of oligosaccharides was performed using PGC cartridges. Cartridges were primed twice with 2 mL of 80% ACN with 0.1% (v/v) TFA and then conditioned twice with 2 mL of water prior to sample loading. After loading the entire reaction mixture, samples were washed five times with 2 mL of water. Elution of oligosaccharides was performed using 2 x 2 mL of 40% ACN with 0.05% (v/v) TFA. Vacuum centrifugation was used to dry the samples to completion.

High performance liquid chromatography-quadrupole time of flight mass spectrometry (HPLC-Q-TOF-MS) analysis

Samples were reconstituted in 100 µL of nanopure water before HPLC-Q-TOF-MS analysis. Analytical separation was carried out using an Agilent 1260 Infinity II HPLC coupled to an Agilent 6530 Accurate-Mass Q-TOF mass spectrometer operated in the positive mode. Chromatographic separation was performed on an analytical PGC column (Hypercarb, 5 µm, 1 x 150 mm, Thermo Scientific). A binary gradient was employed and consisted of solvent A (3% ACN/97% H_2O with 0.1% FA) and solvent B (90% ACN/10% H_2O with 0.1% FA). A 45-min gradient with a flow rate of 0.150 mL/min was used for chromatographic separation: 3-25% B, 0-15 min; 25% B, 15-18 min; 25-99% B, 18-30 min; 99% B, 30-32 min; 99-3% B, 32-34 min; 3% B, 34-45 min. The mass spectrometer was run in positive mode and a reference mass with 922.0098 *m*/*z* was used for internal mass calibration. Drying gas temperature and flow rate were set to 150 ºC and 11 L/min, respectively. Operation voltages for the fragment, skimmer, and octupole 1 RF were 175, 60, and

750 V, respectively. The acquisition rate was set to 0.63 spectra/second. For fragmentation, the linear function, Collision Energy = 1.45*(*m/z*)/100 - 3.5, was employed. Data obtained from the HPLC-QTOF MS were collected using Agilent MassHunter Workstation Data Acquisition version B.06.01. The acquired data were analyzed using Agilent MassHunter Quantitative Analysis version B.06.00. Oligosaccharide compositions were manually identified and annotated using tandem MS data, where neutral mass losses of the monosaccharides were used as basis for assigning monosaccharide class composition. The LC-MS profiles were annotated with the number of monosaccharides per monosaccharide class (Hexose or Hex, Pentose or Pent, 4-*O*-Methylated Glucuronic Acid or GlcA-OMe) involved in the makeup of the identified oligosaccharide. The number of monosaccharides was represented as a subscript. For example, oligosaccharides containing a mixture of monosaccharide classes were labeled as Hex_nPent_m, where *n* represents the number of hexoses and *m* represents the number of pentoses. Thus, a monosaccharide composition of Hex_3Pent indicated the presence of an oligosaccharide composed of 3 hexoses and 1 pentose resulting in an overall degree of polymerization (DP) of 4.

Measurement of similarity between chromatographic profiles

Two methods were employed to examine the similarity between the chromatographic profiles of standard and food polysaccharide samples. Peak coverage determines the percentage of oligosaccharide peaks observed in the food sample relative to the polysaccharide standard. Therefore, the higher the number of matched oligosaccharide peaks between the food and standard polysaccharide LC-MS profiles, the higher the percent peak coverage value for that polysaccharide.

A second approach involves a chemometric technique using the angle cosine method. In this method, the two chromatograms under investigation are treated as vectors of peak areas. The

number of oligosaccharide peaks along with the corresponding area are included in the similarity computation. The angle cosine method was applied to measure the similarity between the two vectors which was calculated using **Equation (2.1):**

$$r^{cos} = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i^2}} \quad \textbf{(2.1)}$$

where x_i and y_i refer to the chromatographic peak areas of oligosaccharide i between the two samples, respectively and n is the number of chromatographic peaks. Oligosaccharides were matched between two chromatograms using retention time values and compound masses. In this manner, a r^{cos} value of 0 indicates that there is no similarity between the two chromatograms while a value of 1 signifies that the two chromatograms are the same.

RESULTS AND DISCUSSION

The method employs oxidative degradation of polysaccharides followed by reduction of product oligosaccharides and purification before HPLC-QTOF MS analysis (**Figure 2.1**). Depolymerization of polysaccharides to oligosaccharides was performed using a metal catalyst, $Fe_2(SO_4)_3$, and an oxidizing agent, H_2O_2, to produce hydroxyl radicals. Oligosaccharides were released and neutralized using NaOH and glacial acetic acid, respectively. The generated oligosaccharides were reduced using $NaBH_4$ to prevent chromatographic anomer separation during analysis. A final cleanup procedure employing solid phase extraction was performed to purify the oligosaccharide fraction. Reduced oligosaccharides were then analyzed by HPLC-QTOF MS.

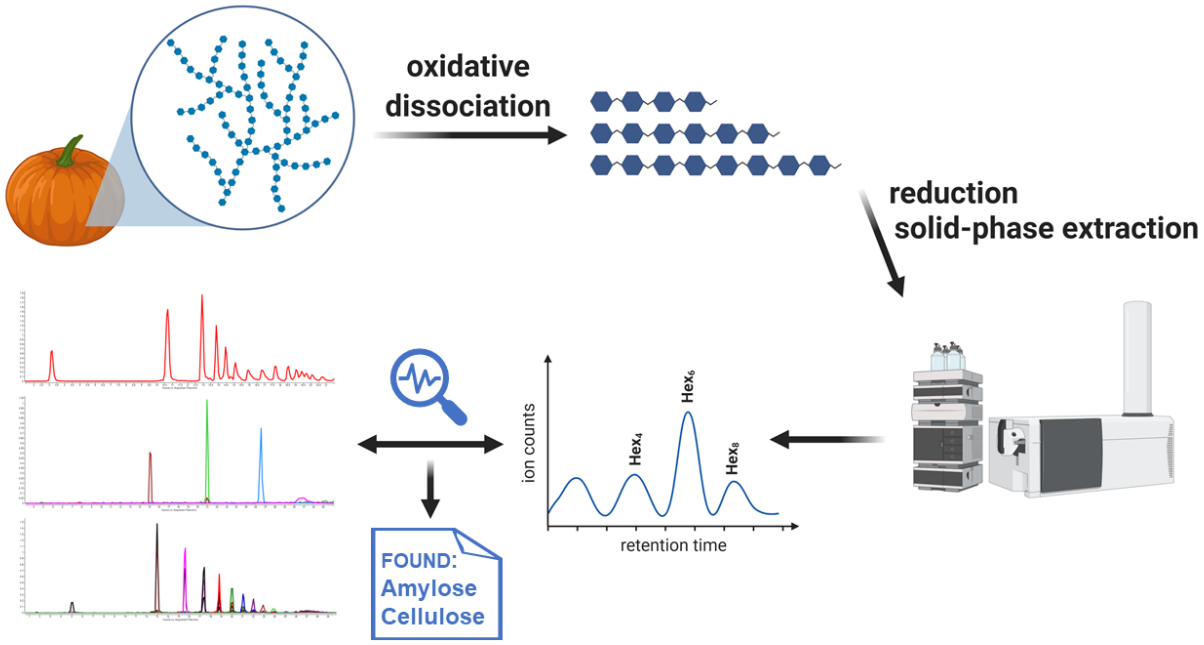


Figure 2.1. Oxidative degradation of polysaccharides generates representative oligosaccharides which were reduced and purified for HPLC-QTOF MS analysis. Oligosaccharides identified in the LC-MS profile were aligned with the oligosaccharide fingerprint library to identify the corresponding parent polysaccharide.

Optimization of reaction conditions

To achieve the optimal reaction conditions, several parameters were optimized including the concentration of $Fe_2(SO_4)_3$ and H_2O_2, pH, reaction time and temperature. A substrate with a diverse monosaccharide and glycosidic linkage composition and a high degree of branching was chosen for the optimization of the reaction conditions. Xyloglucan, a heteropolysaccharide known to contain a β(1→4) glucose backbone with occasional α(1→6) xylose side-chains capped with galactose residues [23] was used for the optimization of the concentrations of $Fe_2(SO_4)_3$ and H_2O_2, pH, reaction time and temperature. The efficacy of the reaction was monitored by examining the total peak area and average DP of the generated xyloglucan oligosaccharides in the chromatogram.

Concentrations of $Fe_2(SO_4)_3$ and H_2O_2 were found to have significant roles on the overall efficiency of the oxidation reaction [24,25]. The optimal $Fe_2(SO_4)_3$ concentration was determined by comparing the total peak areas of the generated xyloglucan oligosaccharides with concentrations of 0.0065, 0.065, 0.65, 1.95, 6.5, and 65 µM. Maximum yield of oligosaccharides was observed at 65 µM of $Fe_2(SO_4)_3$ (**Figure 2.2A**). The optimized $Fe_2(SO_4)_3$ concentration was then used to optimize for the concentration of H_2O_2. The H_2O_2 concentration was varied at concentrations of 0.06, 0.29, 0.58, and 1.16 M. Oligosaccharide yield was highest at 0.29 M of H_2O_2 (**Figure 2.2B**). Correspondingly, the highest average DP of the generated product oligosaccharides was also observed at 0.29 M H_2O_2.

To optimize the buffer pH, a pH range between 2.0 to 12.0 was evaluated with the optimized concentrations of $Fe_2(SO_4)_3$ and H_2O_2. Previous studies on similar reactions revealed a strong pH dependency [26,27]. The optimal pH for efficient oligosaccharide production was determined to be approximately 5.0 as shown in **Figure 2.2C**. At pH 12.0, the average DP increased further. However, the total peak area of oligosaccharides was substantially lower. Under

alkaline conditions, the decline in the progress of the oxidation reaction was attributed to the precipitation of $Fe(OH)_3$ [24].

Reaction time and temperature were also determined to be important parameters in the efficiency of the oxidation reaction. The temperature was varied between 25 ℃ and 100 ℃ in increments of 25 ℃ while time was varied between 0 min and 120 min in increments of 20 min. Optimal yield was obtained at a temperature and time of 100 ℃ and 20 min (**Figure 2.2D**). The highest average DP of product oligosaccharides in the chromatogram was observed at 75 ℃ after 60 min (**Figure 2.2E**). However, the total yield at these conditions was substantially lower. Thus, a temperature of 100 ℃ and time of 20 min was chosen as the optimal condition for effective polysaccharide degradation.

Generation of hydroxyl radicals would yield not only glycosidic cleavage products but also other reaction side products such as oxidatively modified oligosaccharides. These side products, however, were not monitored because they are not used in the fingerprinting analysis. The reaction conditions were optimized to give the maximum yield of unmodified oligosaccharides since these products are more informative for deducing the parent polysaccharide structure.

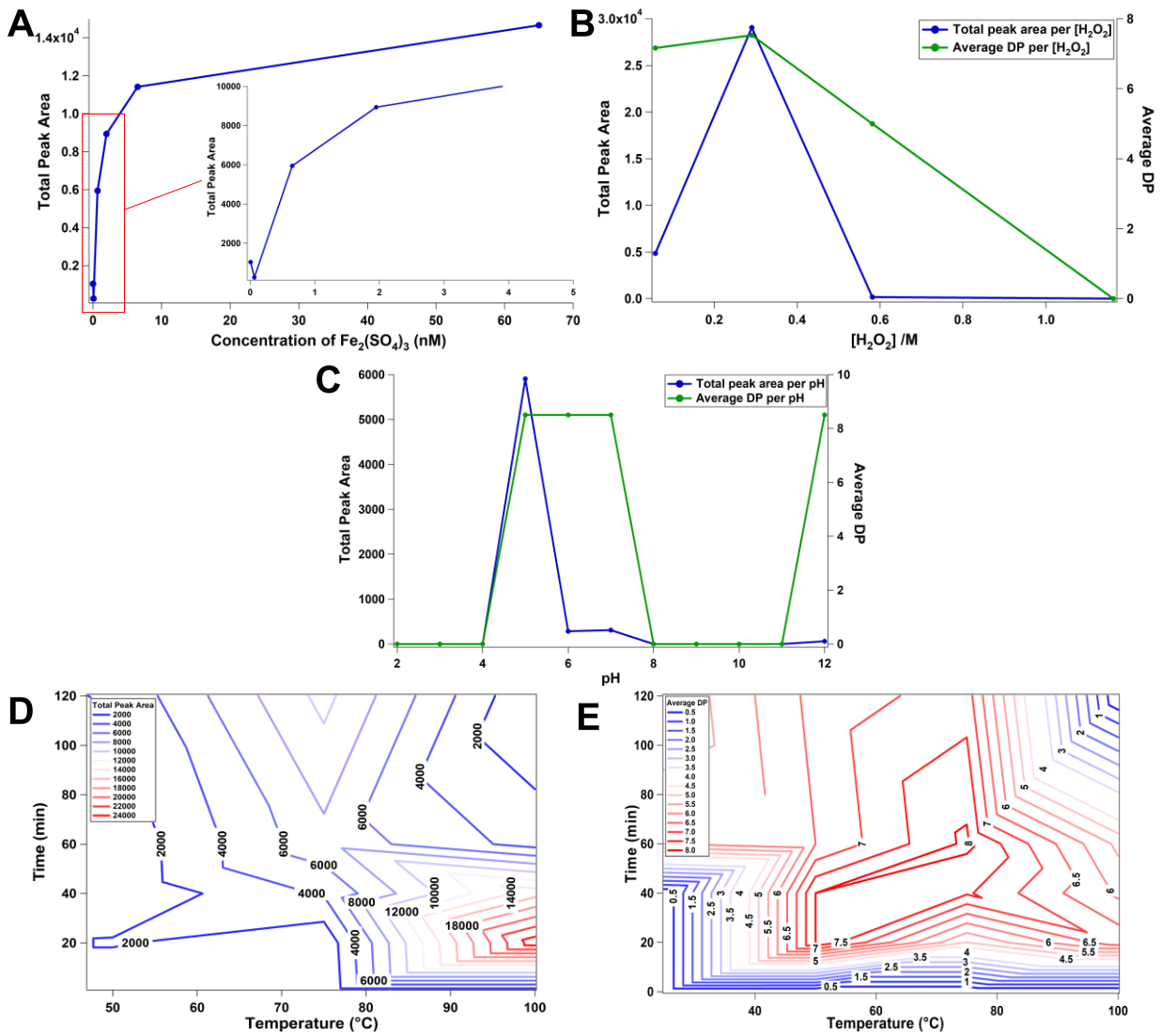


Figure 2.2. Effective polysaccharide degradation was performed by optimizing several reaction parameters including the concentration of $Fe_2(SO_4)_3$ (**A**) and H_2O_2 (**B**), and pH (**C**). Time and temperature combination were also optimized to ensure high total peak area (**D**) and average DP (**E**) of the generated oligosaccharides.

Method validation

The optimized reaction conditions were used to generate a series of oligosaccharides that would fingerprint the corresponding plant polysaccharides. These oligosaccharides generated from commercially available polysaccharide standards were tabulated in **Table 2.1,** which includes their

retention time, monoisotopic mass, oligosaccharide composition, and parent polysaccharide of each identified oligosaccharide. It was observed that each polysaccharide structure generated unique oligosaccharide markers that would be useful for a confident oligosaccharide identification in the sample matrix. Xylan, for example, produced oligosaccharides with methylated glucuronic acid residues, showing that the degradation reaction was able to preserve structural information of the parent polysaccharide. Most isomeric oligosaccharides with different linkage configurations are well-separated using the optimized LC method. PGC has been commonly used to effectively separate isomeric oligosaccharides [22,28,29]. Thus, a Hex_4 oligosaccharide from different parent polysaccharides, such as amylose, cellulose, galactan, and mannan, are all separated by the LC method. This highlights one of the advantages of this method in distinguishing closely related polysaccharide structures such as amylose and cellulose. Both polysaccharides gave distinct oligosaccharide profiles useful for identification. Additionally, isomeric oligosaccharides were also produced in more complex polysaccharides such β-glucan, wherein glucose units were attached as either β(1→4) or β(1→3) linkage, producing multiple isomers for each DP, as high as DP of 6.

Table 2.1. Comprehensive library of oligosaccharides generated from polysaccharide standards using the optimized reaction conditions for oxidative dissociation. Retention time, monoisotopic mass (non-reduced), oligosaccharide composition, and parent polysaccharide information are reported.

Retention Time (mins)	Monoisotopic Mass (Da)	Oligosaccharide Composition	Parent Polysaccharide
3.32	504.169	Hex_3	Amylose
10.67	666.222	Hex_4	Amylose
13.27	828.274	Hex_5	Amylose
14.36	990.327	Hex_6	Amylose
15.06	1152.380	Hex_7	Amylose
15.85	1314.433	Hex_8	Amylose

Retention Time (mins)	Monoisotopic Mass (Da)	Oligosaccharide Composition	Parent Polysaccharide
16.80	1476.486	Hex_9	Amylose
17.76	1638.538	Hex_{10}	Amylose
18.81	1800.591	Hex_{11}	Amylose
19.77	1962.644	Hex_{12}	Amylose
20.40	2124.697	Hex_{13}	Amylose
20.83	2286.750	Hex_{14}	Amylose
21.14	2448.802	Hex_{15}	Amylose
21.62	2610.855	Hex_{16}	Amylose
22.31	2772.908	Hex_{17}	Amylose
23.36	2934.961	Hex_{18}	Amylose
24.27	3097.014	Hex_{19}	Amylose
25.16	3259.066	Hex_{20}	Amylose
25.70	3421.119	Hex_{21}	Amylose
2.97	414.137	Pnt_3	Arabinan
3.78	546.180	Pnt_4	Arabinan
4.83	414.137	Pnt_3	Arabinan
10.25	414.137	Pnt_3	Arabinan
11.18	546.180	Pnt_4	Arabinan
12.02	546.180	Pnt_4	Arabinan
12.68	546.180	Pnt_4	Arabinan
13.28	546.180	Pnt_4	Arabinan
14.61	678.222	Pnt_5	Arabinan
15.40	546.180	Pnt_4	Arabinan
15.52	678.222	Pnt_5	Arabinan
15.76	810.264	Pnt_6	Arabinan
16.16	678.222	Pnt_5	Arabinan
17.35	810.264	Pnt_6	Arabinan
18.56	942.307	Pnt_7	Arabinan
18.88	942.307	Pnt_7	Arabinan
19.35	942.307	Pnt_7	Arabinan
19.87	942.307	Pnt_7	Arabinan

Retention Time (mins)	Monoisotopic Mass (Da)	Oligosaccharide Composition	Parent Polysaccharide
20.26	1074.349	Pnt_8	Arabinan
21.64	1206.391	Pnt_9	Arabinan
22.96	1338.433	Pnt_{10}	Arabinan
23.78	1470.476	Pnt_{11}	Arabinan
24.99	1602.518	Pnt_{12}	Arabinan
26.83	1734.560	Pnt_{13}	Arabinan
1.73	504.169	Hex_3	Arabinogalactan
2.18	474.158	Hex_2Pnt	Arabinogalactan
2.30	504.169	Hex_3	Arabinogalactan
2.85	666.222	Hex_4	Arabinogalactan
3.21	576.190	$HexPnt_3$	Arabinogalactan
3.37	666.222	Hex_4	Arabinogalactan
3.82	606.201	Hex_2Pnt_2	Arabinogalactan
4.00	414.137	Pnt_3	Arabinogalactan
4.68	666.222	Hex_4	Arabinogalactan
4.83	504.169	Hex_3	Arabinogalactan
6.59	828.274	Hex_5	Arabinogalactan
6.95	666.222	Hex_4	Arabinogalactan
7.28	828.274	Hex_5	Arabinogalactan
7.28	636.211	Hex_3Pnt	Arabinogalactan
9.22	474.158	Hex_2Pnt	Arabinogalactan
11.21	474.158	Hex_2Pnt	Arabinogalactan
11.36	990.327	Hex_6	Arabinogalactan
11.50	828.274	Hex_5	Arabinogalactan
11.66	474.158	Hex_2Pnt	Arabinogalactan
12.14	666.222	Hex_4	Arabinogalactan
12.26	636.211	Hex_3Pnt	Arabinogalactan
12.80	828.274	Hex_5	Arabinogalactan
12.97	666.222	Hex_4	Arabinogalactan
13.31	636.211	Hex_3Pnt	Arabinogalactan
13.31	606.201	Hex_2Pnt_2	Arabinogalactan

Retention Time (mins)	Monoisotopic Mass (Da)	Oligosaccharide Composition	Parent Polysaccharide
13.48	828.274	Hex_5	Arabinogalactan
13.63	990.327	Hex_6	Arabinogalactan
14.30	1152.380	Hex_7	Arabinogalactan
15.00	990.327	Hex_6	Arabinogalactan
15.00	960.317	Hex_5Pnt	Arabinogalactan
15.44	990.327	Hex_6	Arabinogalactan
16.25	1152.380	Hex_7	Arabinogalactan
16.37	990.327	Hex_6	Arabinogalactan
16.51	546.180	Pnt_4	Arabinogalactan
16.69	1152.380	Hex_7	Arabinogalactan
17.00	1314.433	Hex_8	Arabinogalactan
17.19	798.264	Hex_4Pnt	Arabinogalactan
17.54	1152.380	Hex_7	Arabinogalactan
18.10	1314.433	Hex_8	Arabinogalactan
18.44	1314.433	Hex_8	Arabinogalactan
18.81	828.274	Hex_5	Arabinogalactan
2.04	414.137	Pnt_3	Arabinoxylan
2.67	414.137	Pnt_3	Arabinoxylan
4.23	414.137	Pnt_3	Arabinoxylan
4.74	546.180	Pnt_4	Arabinoxylan
7.48	546.180	Pnt_4	Arabinoxylan
8.23	546.180	Pnt_4	Arabinoxylan
10.76	546.180	Pnt_4	Arabinoxylan
12.07	546.180	Pnt_4	Arabinoxylan
12.38	546.180	Pnt_4	Arabinoxylan
13.15	546.180	Pnt_4	Arabinoxylan
13.41	678.222	Pnt_5	Arabinoxylan
13.69	678.222	Pnt_5	Arabinoxylan
14.25	678.222	Pnt_5	Arabinoxylan
14.55	678.222	Pnt_5	Arabinoxylan
14.91	678.222	Pnt_5	Arabinoxylan

Retention Time (mins)	Monoisotopic Mass (Da)	Oligosaccharide Composition	Parent Polysaccharide
15.59	678.222	Pnt_5	Arabinoxylan
16.11	678.222	Pnt_5	Arabinoxylan
16.47	414.137	Pnt_3	Arabinoxylan
16.60	810.264	Pnt_6	Arabinoxylan
17.05	678.222	Pnt_5	Arabinoxylan
17.30	810.264	Pnt_6	Arabinoxylan
18.03	678.222	Pnt_5	Arabinoxylan
18.45	546.180	Pnt_4	Arabinoxylan
18.53	678.222	Pnt_5	Arabinoxylan
18.77	810.264	Pnt_6	Arabinoxylan
19.13	546.180	Pnt_4	Arabinoxylan
19.65	810.264	Pnt_6	Arabinoxylan
19.75	546.180	Pnt_4	Arabinoxylan
20.24	810.264	Pnt_6	Arabinoxylan
21.16	810.264	Pnt_6	Arabinoxylan
21.39	810.264	Pnt_6	Arabinoxylan
21.67	942.307	Pnt_7	Arabinoxylan
21.94	810.264	Pnt_6	Arabinoxylan
22.06	678.222	Pnt_5	Arabinoxylan
22.59	942.307	Pnt_7	Arabinoxylan
22.71	810.264	Pnt_6	Arabinoxylan
23.14	810.264	Pnt_6	Arabinoxylan
24.41	678.222	Pnt_5	Arabinoxylan
24.55	942.307	Pnt_7	Arabinoxylan
25.37	942.307	Pnt_7	Arabinoxylan
26.06	810.264	Pnt_6	Arabinoxylan
27.62	1074.349	Pnt_8	Arabinoxylan
28.46	942.307	Pnt_7	Arabinoxylan
3.26	504.169	Hex_3	β-Glucan
4.49	504.169	Hex_3	β-Glucan
7.44	504.169	Hex_3	β-Glucan

Retention Time (mins)	Monoisotopic Mass (Da)	Oligosaccharide Composition	Parent Polysaccharide
10.59	666.222	Hex_4	β-Glucan
11.89	504.169	Hex_3	β-Glucan
12.46	504.169	Hex_3	β-Glucan
13.20	504.169	Hex_3	β-Glucan
14.94	504.169	Hex_3	β-Glucan
16.63	504.169	Hex_3	β-Glucan
17.21	666.222	Hex_4	β-Glucan
17.56	666.222	Hex_4	β-Glucan
22.24	666.222	Hex_4	β-Glucan
22.31	828.274	Hex_5	β-Glucan
25.00	828.274	Hex_5	β-Glucan
25.98	828.274	Hex_5	β-Glucan
27.31	990.327	Hex_6	β-Glucan
31.18	990.327	Hex_6	β-Glucan
14.28	504.169	Hex_3	Cellulose
18.77	666.222	Hex_4	Cellulose
27.33	828.274	Hex_5	Cellulose
13.70	504.169	Hex_3	Curdlan
20.72	666.222	Hex_4	Curdlan
27.96	828.274	Hex_5	Curdlan
2.158	504.169	Hex_3	Galactan
4.96	666.222	Hex_4	Galactan
10.57	828.274	Hex_5	Galactan
12.35	990.327	Hex_6	Galactan
13.39	1152.380	Hex_7	Galactan
14.16	1314.433	Hex_8	Galactan
2.58	504.169	Hex_3	Galactomannan
3.88	666.222	Hex_4	Galactomannan
4.25	666.222	Hex_4	Galactomannan
4.67	666.222	Hex_4	Galactomannan
8.18	666.222	Hex_4	Galactomannan

Retention Time (mins)	Monoisotopic Mass (Da)	Oligosaccharide Composition	Parent Polysaccharide
9.99	828.274	Hex_5	Galactomannan
11.02	828.274	Hex_5	Galactomannan
11.34	828.274	Hex_5	Galactomannan
12.14	666.222	Hex_4	Galactomannan
12.26	828.274	Hex_5	Galactomannan
13.03	990.327	Hex_6	Galactomannan
13.15	828.274	Hex_5	Galactomannan
13.57	990.327	Hex_6	Galactomannan
13.79	990.327	Hex_6	Galactomannan
14.10	990.327	Hex_6	Galactomannan
14.63	1152.380	Hex_7	Galactomannan
14.85	990.327	Hex_6	Galactomannan
15.21	666.222	Hex_4	Galactomannan
15.35	1152.380	Hex_7	Galactomannan
15.90	1314.433	Hex_8	Galactomannan
16.48	990.327	Hex_6	Galactomannan
16.63	1152.380	Hex_7	Galactomannan
16.93	1314.433	Hex_8	Galactomannan
1.92	504.169	Hex_3	Glucomannan
5.62	666.222	Hex_4	Glucomannan
6.55	504.169	Hex_3	Glucomannan
8.73	504.169	Hex_3	Glucomannan
10.62	666.222	Hex_4	Glucomannan
11.45	828.274	Hex_5	Glucomannan
11.60	666.222	Hex_4	Glucomannan
11.84	504.169	Hex_3	Glucomannan
12.27	666.222	Hex_4	Glucomannan
12.92	666.222	Hex_4	Glucomannan
12.93	990.327	Hex_6	Glucomannan
13.20	828.274	Hex_5	Glucomannan
13.56	990.327	Hex_6	Glucomannan

Retention Time (mins)	Monoisotopic Mass (Da)	Oligosaccharide Composition	Parent Polysaccharide
14.02	504.169	Hex_3	Glucomannan
14.31	990.327	Hex_6	Glucomannan
15.02	1152.380	Hex_7	Glucomannan
15.26	828.274	Hex_5	Glucomannan
15.52	990.327	Hex_6	Glucomannan
15.82	1314.433	Hex_8	Glucomannan
15.94	828.274	Hex_5	Glucomannan
16.23	666.222	Hex_4	Glucomannan
16.46	990.327	Hex_6	Glucomannan
16.62	828.274	Hex_5	Glucomannan
16.77	666.222	Hex_4	Glucomannan
17.02	1152.380	Hex_7	Glucomannan
17.14	828.274	Hex_5	Glucomannan
17.30	990.327	Hex_6	Glucomannan
17.62	828.274	Hex_5	Glucomannan
17.74	990.327	Hex_6	Glucomannan
17.86	828.274	Hex_5	Glucomannan
18.01	666.222	Hex_4	Glucomannan
18.26	828.274	Hex_5	Glucomannan
18.37	1152.380	Hex_7	Glucomannan
18.51	990.327	Hex_6	Glucomannan
18.79	1152.380	Hex_7	Glucomannan
19.06	828.274	Hex_5	Glucomannan
19.14	990.327	Hex_6	Glucomannan
19.24	1152.380	Hex_7	Glucomannan
19.40	990.327	Hex_6	Glucomannan
20.03	1152.380	Hex_7	Glucomannan
20.31	1152.380	Hex_7	Glucomannan
20.83	1314.433	Hex_8	Glucomannan
21.11	666.222	Hex_4	Glucomannan
21.54	1152.380	Hex_7	Glucomannan

Retention Time (mins)	Monoisotopic Mass (Da)	Oligosaccharide Composition	Parent Polysaccharide
21.85	828.274	Hex_5	Glucomannan
21.99	1152.380	Hex_7	Glucomannan
22.27	828.274	Hex_5	Glucomannan
22.76	990.327	Hex_6	Glucomannan
23.22	1152.380	Hex_7	Glucomannan
23.88	1152.380	Hex_7	Glucomannan
24.72	990.327	Hex_6	Glucomannan
25.42	828.274	Hex_5	Glucomannan
25.77	990.327	Hex_6	Glucomannan
27.00	1314.433	Hex_8	Glucomannan
27.68	1476.486	Hex_9	Glucomannan
2.04	504.169	Hex_3	Lichenan
2.88	666.222	Hex_4	Lichenan
3.06	504.169	Hex_3	Lichenan
3.34	504.169	Hex_3	Lichenan
5.85	504.169	Hex_3	Lichenan
6.14	828.274	Hex_5	Lichenan
7.59	828.274	Hex_5	Lichenan
9.07	666.222	Hex_4	Lichenan
10.10	990.327	Hex_6	Lichenan
10.78	990.327	Hex_6	Lichenan
11.08	828.274	Hex_5	Lichenan
12.13	1152.380	Hex_7	Lichenan
12.59	504.169	Hex_3	Lichenan
12.89	990.327	Hex_6	Lichenan
13.42	828.274	Hex_5	Lichenan
14.29	828.274	Hex_5	Lichenan
14.75	990.327	Hex_6	Lichenan
15.02	504.169	Hex_3	Lichenan
15.34	990.327	Hex_6	Lichenan
16.53	828.274	Hex_5	Lichenan

Retention Time (mins)	Monoisotopic Mass (Da)	Oligosaccharide Composition	Parent Polysaccharide
16.62	1152.380	Hex_7	Lichenan
17.23	990.327	Hex_6	Lichenan
17.62	990.327	Hex_6	Lichenan
18.10	1152.380	Hex_7	Lichenan
18.52	1152.380	Hex_7	Lichenan
18.96	1314.433	Hex_8	Lichenan
20.39	1152.380	Hex_7	Lichenan
20.99	1314.433	Hex_8	Lichenan
21.45	1476.486	Hex_9	Lichenan
22.22	666.222	Hex_4	Lichenan
24.88	828.274	Hex_5	Lichenan
25.82	828.274	Hex_5	Lichenan
27.31	990.327	Hex_6	Lichenan
31.06	990.327	Hex_6	Lichenan
1.66	504.169	Hex_3	Mannan
3.33	666.222	Hex_4	Mannan
9.50	828.274	Hex_5	Mannan
13.03	990.327	Hex_6	Mannan
14.27	1152.380	Hex_7	Mannan
15.22	1314.433	Hex_8	Mannan
15.98	1476.486	Hex_9	Mannan
16.62	1638.538	Hex_{10}	Mannan
17.25	1800.591	Hex_{11}	Mannan
17.69	1962.644	Hex_{12}	Mannan
18.24	2124.697	Hex_{13}	Mannan
18.70	2286.750	Hex_{14}	Mannan
19.29	2448.802	Hex_{15}	Mannan
19.84	2610.855	Hex_{16}	Mannan
10.14	414.137	Pnt_3	Xylan
13.31	604.190	$Pnt_3GcaOMe$	Xylan
14.29	604.190	$Pnt_3GcaOMe$	Xylan

Retention Time (mins)	Monoisotopic Mass (Da)	Oligosaccharide Composition	Parent Polysaccharide
14.72	604.190	$Pnt_3GcaOMe$	Xylan
16.52	546.180	Pnt_4	Xylan
17.11	604.190	$Pnt_3GcaOMe$	Xylan
18.36	736.232	$Pnt_4GcaOMe$	Xylan
18.80	736.232	$Pnt_4GcaOMe$	Xylan
20.16	678.222	Pnt_5	Xylan
20.45	736.232	$Pnt_4GcaOMe$	Xylan
21.41	736.232	$Pnt_4GcaOMe$	Xylan
21.56	868.274	$Pnt_5GcaOMe$	Xylan
22.41	868.274	$Pnt_5GcaOMe$	Xylan
23.51	810.264	Pnt_6	Xylan
23.96	868.274	$Pnt_5GcaOMe$	Xylan
24.68	868.274	$Pnt_5GcaOMe$	Xylan
24.85	1000.317	$Pnt_6GcaOMe$	Xylan
26.48	1000.317	$Pnt_6GcaOMe$	Xylan
27.31	942.307	Pnt_7	Xylan
27.65	1000.317	$Pnt_6GcaOMe$	Xylan
27.82	1132.359	$Pnt_7GcaOMe$	Xylan
29.11	1074.349	Pnt_8	Xylan
29.40	1132.359	$Pnt_7GcaOMe$	Xylan
5.87	474.158	Hex_2Pnt	Xyloglucan
7.52	474.158	Hex_2Pnt	Xyloglucan
11.88	474.158	Hex_2Pnt	Xyloglucan
12.33	606.201	Hex_2Pnt_2	Xyloglucan
13.29	636.211	Hex_3Pnt	Xyloglucan
14.06	636.211	Hex_3Pnt	Xyloglucan
14.83	606.201	Hex_2Pnt_2	Xyloglucan
15.01	504.169	Hex_3	Xyloglucan
16.29	768.253	Hex_3Pnt_2	Xyloglucan
17.04	636.211	Hex_3Pnt	Xyloglucan
17.16	930.306	Hex_4Pnt_2	Xyloglucan

Retention Time (mins)	Monoisotopic Mass (Da)	Oligosaccharide Composition	Parent Polysaccharide
17.48	606.201	Hex_2Pnt_2	Xyloglucan
17.96	606.201	Hex_2Pnt_2	Xyloglucan
18.06	768.253	Hex_3Pnt_2	Xyloglucan
18.16	798.264	Hex_4Pnt	Xyloglucan
18.29	768.253	Hex_3Pnt_2	Xyloglucan
18.77	636.211	Hex_3Pnt	Xyloglucan
19.47	930.306	Hex_4Pnt_2	Xyloglucan
20.00	900.296	Hex_3Pnt_3	Xyloglucan
20.12	930.306	Hex_4Pnt_2	Xyloglucan
20.36	1092.359	Hex_5Pnt_2	Xyloglucan
20.66	930.306	Hex_4Pnt_2	Xyloglucan
21.01	930.306	Hex_4Pnt_2	Xyloglucan
21.47	1092.359	Hex_5Pnt_2	Xyloglucan
21.17	666.222	Hex_4	Xyloglucan
21.80	900.296	Hex_3Pnt_3	Xyloglucan
22.09	1062.349	Hex_4Pnt_3	Xyloglucan
22.40	1062.349	Hex_4Pnt_3	Xyloglucan
22.74	1224.401	Hex_5Pnt_3	Xyloglucan
23.77	798.264	Hex_4Pnt	Xyloglucan
24.11	1092.359	Hex_5Pnt_2	Xyloglucan
24.67	1254.412	Hex_6Pnt_2	Xyloglucan
24.85	930.306	Hex_4Pnt_2	Xyloglucan
25.20	1092.359	Hex_5Pnt_2	Xyloglucan
25.37	798.264	Hex_4Pnt	Xyloglucan
25.73	1092.359	Hex_5Pnt_2	Xyloglucan
26.29	1092.359	Hex_5Pnt_2	Xyloglucan
26.81	1224.401	Hex_5Pnt_3	Xyloglucan
26.99	930.306	Hex_4Pnt_2	Xyloglucan
27.50	1224.401	Hex_5Pnt_3	Xyloglucan
28.00	1386.454	Hex_6Pnt_3	Xyloglucan
28.72	1062.349	Hex_4Pnt_3	Xyloglucan

Retention Time (mins)	Monoisotopic Mass (Da)	Oligosaccharide Composition	Parent Polysaccharide
29.06	1062.349	Hex_4Pnt_3	Xyloglucan

A validation step was performed to confirm the concept of fingerprinting polysaccharides using diagnostic oligosaccharides. The method was validated using a synthetic mixture of polysaccharide standards including arabinoxylan, xyloglucan, and amylopectin. The polysaccharides, mixed at equivalent ratios by mass, were reacted using the optimized conditions for oxidative degradation to produce representative oligosaccharides. The LC-MS oligosaccharide profile of a mixture of the three polysaccharides is shown in **Figure 2.3D**. Peak annotation and matching were performed using the individual LC-MS profiles for arabinoxylan, xyloglucan, and amylopectin as shown in **Figure 2.3A-C**, respectively.

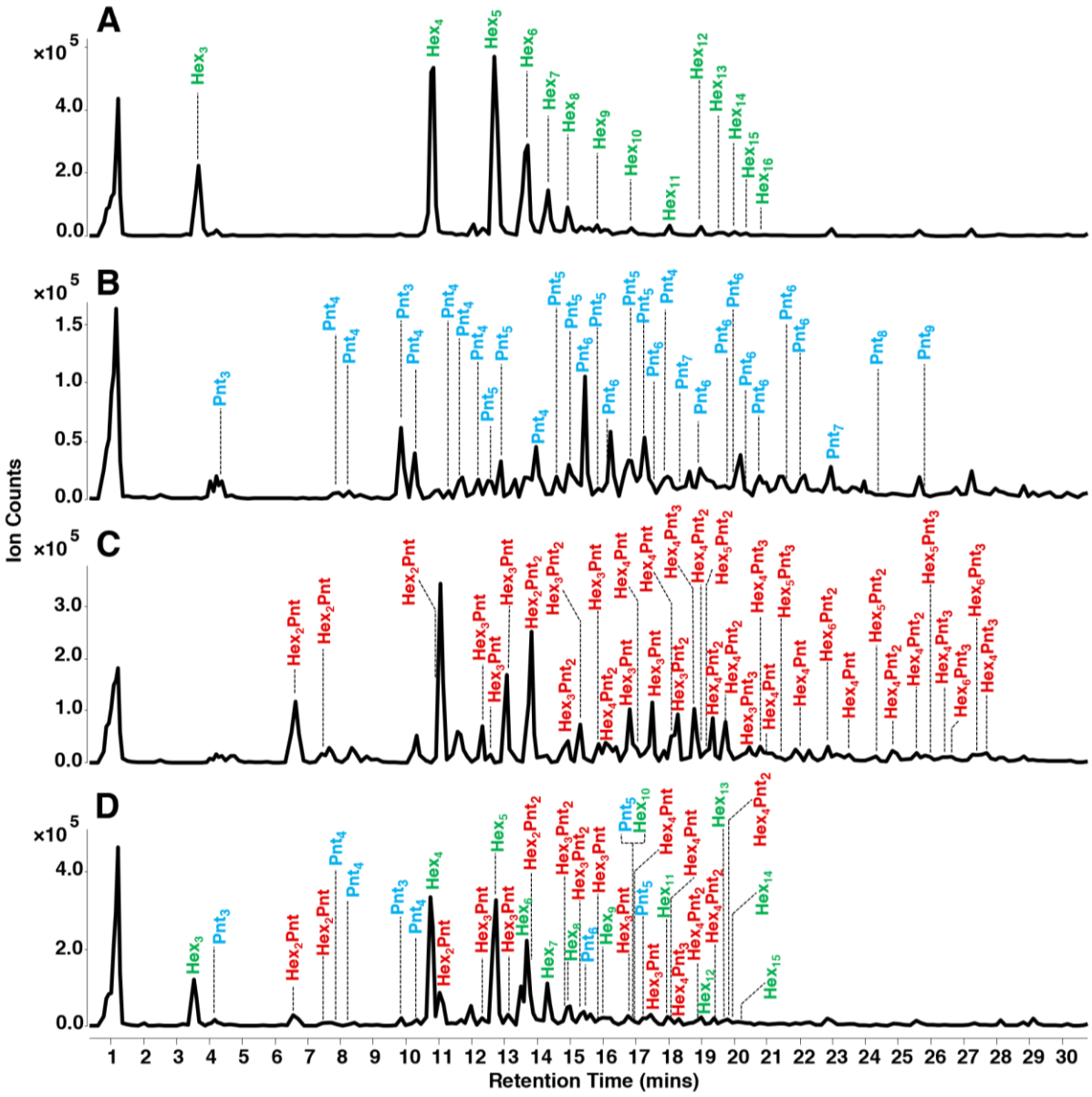


Figure 2.3. Validation of the polysaccharide fingerprinting method was performed using a mixture of three polysaccharide standards. Annotated base-peak chromatograms of the oligosaccharide profiles of amylopectin (**A**), arabinoxylan (**B**), xyloglucan (**C**), and the mixture (**D**) are illustrated. Using oligosaccharide fingerprints, all three polysaccharides were successfully identified in the mixture.

LC-MS profiles of the individual standard polysaccharides were compared with the LC-MS profile of the mixture using the peak coverage method. The peak coverage value of a

polysaccharide represents the percentage of oligosaccharide peaks observed from a mixture in relation to the oligosaccharide peaks observed from a pure polysaccharide solution. When compared to the pure sample, oligosaccharides in the mixture from amylopectin had a peak coverage value of 93%, while xyloglucan and arabinoxylan had 45% and 25%, respectively. The discrepancy in the peak coverage values was ascribed to the wide distribution of molecular weights of the manufactured pure polysaccharides. There was an inherent difference in the molar ratio of the polysaccharides introduced into the oxidative degradation reaction. For this reason, peak coverage values cannot be used for quantitation. Rather, it only provides a metric of confidence for an identification. Polysaccharide quantitation is still an on-going research endeavor in our laboratory. However, the three polysaccharides in the mixture were successfully identified using oligosaccharide fingerprint information, which included monosaccharide composition, retention time, and monoisotopic mass. Thus, employing the oligosaccharide fingerprint library from pure polysaccharides for the identification of unknown polysaccharide compositions in a mixture was successfully validated.

Polysaccharide fingerprinting

The capabilities of the polysaccharide fingerprinting method were probed by analyzing unknown polysaccharide compositions of various foods. The optimized oxidative degradation method for polysaccharides was applied to food samples to generate representative oligosaccharides. Oligosaccharide fingerprints from food LC-MS profiles were matched with the library of oligosaccharide fingerprints generated from standard polysaccharides to confirm the corresponding polysaccharide composition. For example, if a Hex_6 at a retention time of 14.36 min

was observed in the food LC-MS profile, it was inferred using the fingerprinting library that the Hex_6 was generated from amylose. Retention-time shifts were corrected during peak alignment and library matching. Specifically, amylose oligosaccharides were used as retention time standards which were run on the instrument for every 12 sample injections. These served as retention-time markers for the retention-time correction.

In the current version of the oligosaccharide library, most polysaccharides found in plants were included with the exemption of fructans, e.g. inulin and levan, and pectins, e.g. homogalacturonan and rhamnogalacturonan I. Pectins are one of the most abundant classes of plant polysaccharides present in the cell wall matrix, as well as the most complex ones. Fructans are also important plant storage carbohydrates in some plants. We are still optimizing the reaction conditions for these polysaccharides to get sufficient oligosaccharide signal from them. These will be included in the future reports. Additionally, we have added other non-plant polysaccharides in the library, such as curdlan and lichenan. Curdlan is mostly sourced from bacteria while lichenan is primarily sourced from lichens. These were included in the current polysaccharide library to provide more diverse oligosaccharides with both β(1→3)Glc and β(1→4)Glc linkages.

The reproducibility of the oxidative degradation reaction was determined using a food sample. Experimental triplicates of butternut squash were analyzed to ensure that the method generates reproducible LC-MS profiles and polysaccharide compositions. For each identified oligosaccharide, the retention time, oligosaccharide composition, and polysaccharide of origin are shown in **Table 2.2**. Average relative abundances and their corresponding standard deviations were also tabulated. Overall, standard deviation values were less than 3% for relative abundances greater than 1%, demonstrating reproducibility of the method for oxidative degradation of polysaccharides. Moreover, the polysaccharide composition of butternut squash correctly

identified amylose and cellulose, which were in good agreement with literature [30]. Oligosaccharide peaks that were not identified using the library were binned until future assignments.

Table 2.2. Oxidative degradation method reproducibility was performed with experimental replicates (n = 3) of whole butternut squash sample. Average values are reported with corresponding standard deviation.

RT (min)	monosaccharide composition	relative abundance (%)	polysaccharide of origin
10.93 ± 0.01	Hex_4	17 ± 2	amylose
12.85 ± 0.06	Hex_5	14 ± 1	amylose
13.85 ± 0.02	Hex_6	9 ± 0	amylose
14.52 ± 0.02	Hex_7	7 ± 3	amylose
3.71 ± 0.03	Hex_3	5 ± 1	amylose
10.93 ± 0.01	Hex_3Pent	4 ± 1	*
12.89 ± 0.01	Hex_4Pent	4 ± 0	*
14.52 ± 0.02	Hex_8	4 ± 1	amylose
13.85 ± 0.02	Hex_5Pent	3 ± 0	*
3.68 ± 0.03	Hex_2Pent	2 ± 0	*
25.49 ± 0.02	Hex_5	2 ± 3	cellulose
19.82 ± 0.07	Hex_4	2 ± 3	cellulose
14.52 ± 0.02	Hex_6Pent	2 ± 1	*
15.2 ± 0.05	Hex_7Pent	2 ± 2	*
20.22 ± 0.01	Hex_{13}	1 ± 1	amylose
18.42 ± 0.04	Hex_{11}	1 ± 0	amylose
11.92 ± 0.03	Hex_3 #	1 ± 1	*
16.16 ± 0.04	Hex_9	1 ± 0	amylose
19.45 ± 0.06	Hex_{12}	1 ± 0	amylose

20.67 ± 0.1	Hex_{14}	1 ± 1	amylose
21.63 ± 0.004	Hex_{16}	1 ± 1	amylose

** Not present in the oligosaccharide fingerprinting library*

[#]Non-reducing oligosaccharide

In addition to identification of the polysaccharides, an estimation of relative abundances was performed using similarity calculations between the LC-MS profiles of food and standard polysaccharides, specifically, the peak coverage and angle cosine methods. Identified polysaccharides that have high peak coverage values encompassed a greater number of oligosaccharide matches to the corresponding LC-MS profiles of standard polysaccharides. While peak coverage was an adequate method for running a quick measure of similarity by peak count, it did not consider peak areas. Thus, a chemometrics approach, employing the angle cosine method, was additionally used to measure similarity by treating the chromatograms from food and standard polysaccharides as vectors. The angle cosine method calculates r^{cos} values (similarity indices) using **Equation 2.1**, where a value of 1 indicates high similarity and a value of 0 indicates low similarity between the two chromatograms. The calculated similarity indices along with the percent peak coverage values between the LC-MS profiles of food and standard polysaccharides are summarized in **Table 2.3**. These two metrics are expected to correlate but only to some extent since the calculations for each are very distinct from each other. While the peak coverage value is merely a peak counting metric, the cosine similarity considers the relative abundances of the peaks identified in both samples being compared.

An estimation of the abundance of the identified polysaccharides was performed using output values from the angle cosine method [31,32]. Similarity indices were used to estimate which polysaccharides have the highest contribution to the polysaccharide composition of different food

samples. The LC-MS profiles of identified polysaccharides that have high similarity indices resembled the profiles from standard polysaccharides. Therefore, such polysaccharides could have higher abundances in the sample relative to others. Validation of the concept was performed by evaluating of how well similarity indices conformed with the abundances of polysaccharides reported in literature as discussed below.

Table 2.3. Similarity metrics between the LC-MS profiles of food and standard polysaccharides were examined for different food samples. Non-percentage values are similarity indices from the angle cosine method while the percentage values are peak coverage calculations.

Standard Polysaccharides	Coconut flesh	Yellow corn meal	Jackfruit flesh	Guava flesh	Yam leaves	Bok choy leaves	Wheat grass	Whole grain oat	Horseradish root	Spent coffee grounds
Amylose	*	0.947, 95%	0.976, 30%	0.823, 63%	0.978, 79%	*	0.893, 47%	0.863, 63%	0.947, 32%	0.995, 79%
Mannan	0.494, 14%	*	*	*	*	0.472, 14%	*	0.349, 7%	*	*
Cellulose	0.011, 33%	*	0.705, 67%	0.567, 100%	0.239, 67%	0.564, 67%	0.593, 100%	*	0.431, 100%	0.011, 33%
Curdlan	*	*	*	*	0.568, 33%	*	*	*	0.568, 33%	*
Lichenan	0.014, 3%	*	*	*	*	*	*	*	*	*
β-Glucan	0.004, 4%	*	0.178, 13%	*	0.435, 22%	*	*	*	0.555, 9%	0.281, 22%
Xylan	*	*	*	0.500, 4%	*	*	0.500, 4%	*	0.950, 17%	*
Arabinoxylan	*	*	0.032, 2%	0.485, 4%	*	*	*	*	*	*
Galactomannan	0.586, 33%	*	0.381, 7%	*	0.202, 7%	0.188, 15%	*	*	*	0.233, 7%
Glucomannan	0.068, 2%	*	0.165, 2%	0.165, 2%	*	0.103, 3%	*	*	*	*
Arabinogalactan	0.108, 5%	*	*	*	*	*	*	*	*	0.319, 2%
Arabinan	0.006, 4%	*	*	*	0.006, 4%	*	0.006, 4%	0.006, 4%	0.245, 13%	*
Xyloglucan	*	*	*	0.658, 2%	*	*	*	*	*	*

**Polysaccharide not present.*

Food polysaccharide composition analysis

Oligosaccharide profiles of food samples including coconut flesh, yellow corn meal, jackfruit flesh, guava flesh, yam leaves, bok choy leaves, wheat grass, whole grain oatmeal cereal, horseradish root, and spent coffee grounds are shown in **Figure 2.4A-J**. The results of the identified polysaccharides from food samples were compared to the profiles found in literature. However, literature values typically did not provide deep coverage in the polysaccharide analysis of most foods. In such cases, the reported monosaccharide composition analyses were compared with our findings from the polysaccharide fingerprinting method.

The LC-MS profile of the copra of coconut (*Cocos nucifera*) was determined to be composed of galactomannan, cellulose, mannan, arabinogalactan, β-glucan, arabinan, lichenan, and glucomannan (**Figure 2.4A**). In literature, it was reported that the coconut flesh was composed of mannan-based polysaccharides, which were the highest in abundance [33]. From our results, similarity indices from the angle cosine method for galactomannan and mannan were 0.586 and 0.494, respectively. Thus, the similarity indices were most likely directly proportional to the true abundance of the observed polysaccharides. Additionally, our method was capable of differentially identifying mannan, glucomannan, and galactomannan polysaccharides. Based on the results, it was evident that coconut flesh was mainly composed of mannan and galactomannan polysaccharides. Yellow corn meal (*Zea mays*), which is largely composed of corn starch, was expected to contain amylose polysaccharides [34]. Similarly, the results indicated that yellow corn meal was mainly composed of amylose as shown in **Figure 2.4B**. The peak coverage was 95% with a similarity index of 0.947, indicating a high abundance of amylose.

Jackfruit flesh (*Artocarpus heterophyllus*) polysaccharides were determined to be composed of amylose, cellulose, β-glucan, galactomannan, arabinoxylan, and glucomannan

(**Figure 2.4C**). Based on literature, jackfruit flesh was mainly composed of glucose, arabinose, and galactose monosaccharide components [35,36]. While the literature results did not include polysaccharide compositions, the overall polysaccharide composition from the fingerprinting method matched with the reported monosaccharide compositions. Both amylose and cellulose were comparable to the standard profiles as represented by similarity indices of 0.976 and 0.705, respectively. Previous reports indicated a high abundance of glucose, [35,36] which could be attributed to amylose and cellulose. Moreover, several Hex_nPent_m oligosaccharides were also identified. Based on the reported monosaccharide composition, [35,36] arabinose could be one of the potential components of the Hex_nPent_m oligosaccharides.

Guava flesh (*Psidium guajava*) was composed of amylose, cellulose, arabinoxylan, xyloglucan, xylan, and glucomannan. A previous report found that guava flesh was primarily composed of glucose, xylose, and arabinose constituents, [37] confirming the results from the fingerprinting method (**Figure 2.4D**). Here, cellulose and amylose were the large contributors of the glucose content with similarity indices of 0.823 and 0.567, respectively. Additionally, one of the polysaccharides isolated from guava flesh in a previous study was characterized to contain a combination of 3-linked arabinose, 5-linked arabinose, 2,3,5-linked arabinose backbone with occasional glucose branching [38]. Thus, the binned Hex_nPent_m peaks could potentially correspond to the presence of a 'glucoarabinan' type polysaccharide.

Polysaccharides from yam leaves (*Dioscorea sp*.) are composed of β-glucan, cellulose, curdlan, galactomannan, amylose, and arabinan, as shown in **Figure 2.4E**. There have been no report to our best knowledge of polysaccharides in yam leaves, however several studies have reported the presence of mannan in yam [39,40]. Using our method, galactomannan was found with a peak coverage of 7% and a similarity index of 0.202. Based on this observation, galactomannan

was present at a low abundance in yam leaves. Moreover, polysaccharide analysis of other types of yam have found β(1→3) linked glucose residues, [41] which was consistent with the presence of β-glucan and curdlan polysaccharides with similarity indices of 0.435 and 0.568, respectively.

The diverse composition of bok choy leaves (*Brassica rapa*) included cellulose, mannan, galactomannan, and glucomannan (**Figure 2.4F**). Previous neutral monosaccharide analysis indicated a high abundance of glucose, galactose, and mannose residues [42]. Therefore, it was consistent with the presence of cellulose, mannan, glucomannan, and galactomannan. The similarity indices for cellulose, mannan, glucomannan, and galactomannan were 0.564, 0.472, 0.103, and 0.188, respectively. These indices indicate that cellulose and mannan contributed to the high concentration of glucose and mannose. A significant amount of arabinose was previously reported, [42] which could be a possible component of the observed Hex_nPent_m oligosaccharides.

The polysaccharide composition of wheat grass (*Triticum sp.*) included cellulose, xylan, arabinan, and amylose (**Figure 2.4G**). Previous studies reported xylan, arabinan, and β-glucan as significant polysaccharide components of wheat grass [43,44]. Literature from hydrolysis experiments determined high abundance of glucose, xylose, and arabinose residues [45]. The similarity indices of amylose, cellulose, xylan, and arabinan were 0.893, 0.593, 0.500, and 0.006, respectively. Thus, the largest contribution to the overall polysaccharide concentration originated from amylose, cellulose, and xylan. The results were similarly consistent with the previously reported compositions.

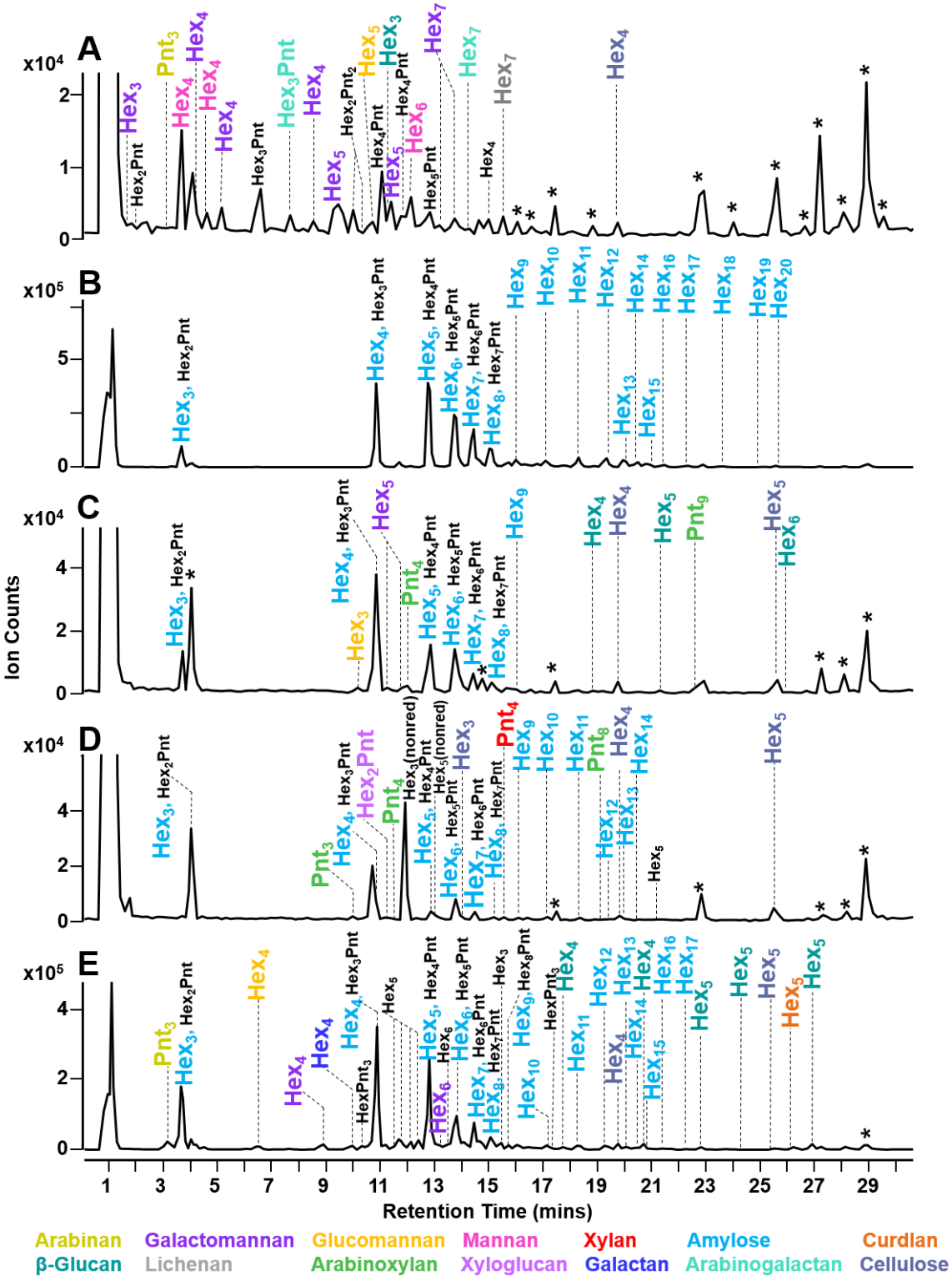
Whole grain oat (*Avena sativa*) cereal was composed of amylose, mannan, and arabinan as shown in **Figure 2.4H**. Several reports indicated the presence of amylose, [46] arabinan, [47] and β-glucans [48,49]. Using the fingerprinting method, amylose and mannan had the highest similarity indices, with 0.863 and 0.349, respectively. The results indicate that whole grain oat is largely

abundant in amylose. Polysaccharides from horseradish (*Armoracia rusticana*) root included arabinan, β-glucan, curdlan, cellulose, xylan, and amylose (**Figure 2.4I**). In literature, cellulose and starch were the most notable components of horseradish roots [50]. Results from the polysaccharide fingerprinting method indicated a predominant presence of amylose and cellulose with similarity indices of 0.947 and 0.431, respectively. The percent peak coverage values were 32% for amylose and 100% for cellulose.

The polysaccharide fingerprinting method was also applied to spent coffee grounds (*Coffea arabica*) (**Figure 2.4J**). Spent coffee grounds were composed of amylose, cellulose, β-glucan, galactomannan, and arabinogalactan. Monosaccharide composition analysis from literature indicated that glucose, galactose, and mannose were major residues present [51]. These results were in agreement with the corresponding polysaccharides from the fingerprinting method. The angle cosine similarity values for amylose, β-glucan, and arabinogalactan were 0.995, 0.568, and 0.319, respectively. Here, amylose and β-glucan are likely to be the more abundant polysaccharide structures, which is comparable to results from the reported monosaccharide analysis [51]. The percent peak coverage values for amylose, β-glucan, and arabinogalactan were 79%, 33%, and 2%, respectively.

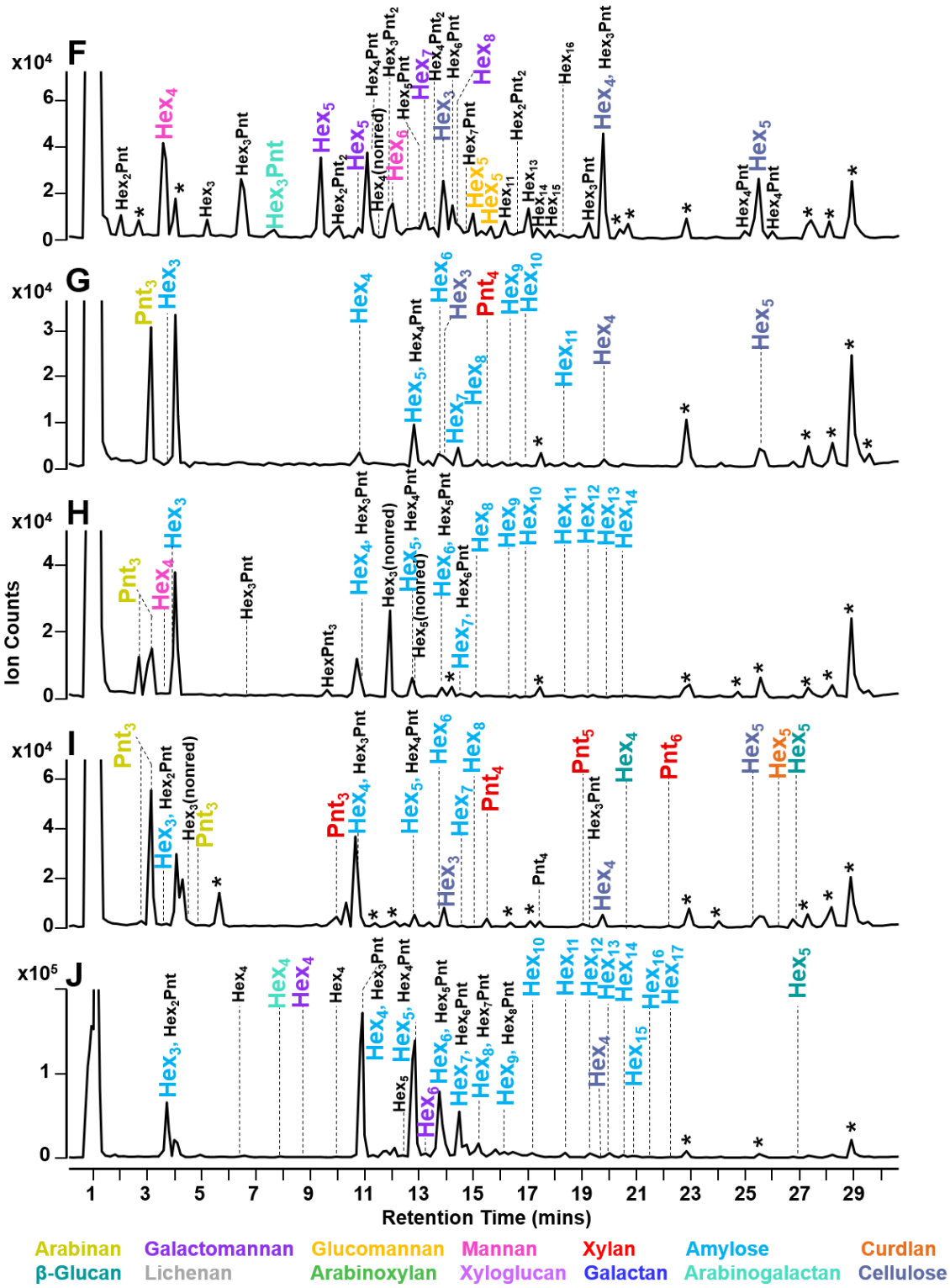
In this set of samples, amylose and cellulose were the two most common polysaccharides identified. The results were expected as these samples were all plant-derived and some of them were rich in starch, as the case for yellow corn meal, while some were expected to be high in cellulose, such as wheat grass. Although curdlan and lichenan are both non-plant polysaccharides, they were detected in few samples but scored low in terms of peak coverage and peak area similarity. One reason might be that these oligosaccharides were from another mixed-linkage glucans similar to β-glucan, curdlan, and/or lichenan. These plant hemicelluloses can have a wide

range of variety in terms of structure, and some of these polysaccharides may have shared some structural moieties. Overall, the analysis has shown the scarcity of information in literature regarding polysaccharide compositions in food. While most food composition analysis include monosaccharide, and some linkage information, very little is known regarding the intact polysaccharide structures. The method presented in this manuscript reveals up a more comprehensive view of carbohydrates in food.



Oligosaccharide not present in the library

∗Non-oligosaccharides



Oligosaccharide not present in the library

*Non-oligosaccharides

Figure 2.4. Food polysaccharide fingerprinting was performed with coconut flesh (**A**), yellow corn meal (**B**), jackfruit flesh (**C**), guava flesh (**D**), yam leaves (**E**), bok choy leaves (**F**),wheat grass (**G**), whole grain oat cereal (**H**), horseradish roots (**I**), and spent coffee grounds (**J**). Corresponding polysaccharides are represented in a color-coded legend. Annotations of co-eluting peaks are separated by a comma. Oligosaccharides that are not present in the library were binned until future assignments. Non-oligosaccharide peaks were denoted with (*****).

CONCLUSIONS

We developed a method for determining the polysaccharide composition of food based on polysaccharide degradation and oligosaccharide fingerprints generated from polysaccharide standards. This method represented a substantial improvement to the slow and stepwise methods for polysaccharide analysis. The oligosaccharide fingerprinting method was validated using a synthetic mixture of standard polysaccharides comprising of xyloglucan, amylose, and arabinoxylan. Method reproducibility was confirmed with experimental triplicates of butternut squash samples, where the overall standard deviation values were calculated to be less than 3% for oligosaccharides with relative abundances greater than 1%. Successful polysaccharide composition identification was performed for ten various food samples. The identified polysaccharide list was validated by comparison with the known compositions in literature. Similarity indices using the angle cosine method was proven to be consistent with previously reported polysaccharide compositions of food samples and demonstrated to be an effective measure of similarity between pure standard and food polysaccharide LC-MS profiles.

Conventional methods for polysaccharide analysis mostly rely on monosaccharide and linkage information to predict the polysaccharide structures in the sample, which often results in several predicted candidates of the parent polysaccharide structures [52]. In contrast to previous techniques, the presented method is capable of differentially identifying polysaccharides in food matrices such as glucose polysaccharides including amylose, cellulose, curdlan, lichenan, and β-

glucan, which would otherwise be rendered indistinguishable from monosaccharide composition data.

Extending the current polysaccharide fingerprinting method for quantitation of polysaccharides would require an orthogonal tool to measure the concentration of identified polysaccharides. In combination with the polysaccharide fingerprinting method, quantitation of polysaccharides using an LC-MS platform is currently being developed and will be the topic of future reports. The presented optimized oxidative method for polysaccharide degradation and the comprehensive library of oligosaccharide fingerprints will allow for a more targeted and rapid workflow for profiling polysaccharides.

REFERENCES

(1) Singdevsachan, S. K.; Auroshree, P.; Mishra, J.; Baliyarsingh, B.; Tayung, K.; Thatoi, H. Mushroom Polysaccharides as Potential Prebiotics with Their Antitumor and Immunomodulating Properties: A Review. *Bioact. Carbohydrates Diet. Fibre* **2016**, *7* (1), 1–14. https://doi.org/10.1016/j.bcdf.2015.11.001.

(2) Kothari, D.; Patel, S.; Kim, S. K. Anticancer and Other Therapeutic Relevance of Mushroom Polysaccharides: A Holistic Appraisal. *Biomed. Pharmacother.* **2018**, *105* (May), 377–394. https://doi.org/10.1016/j.biopha.2018.05.138.

(3) Gross, K. C.; Wallner, S. J. Degradation of Cell Wall Polysaccharides during Tomato Fruit Ripening'. **1979**, 117–120.

(4) Yashoda, H. M.; Prabha, T. N.; Tharanathan, R. N. Mango Ripening - Chemical and Structural Characterization of Pectic and Hemicellulosic Polysaccharides. *Carbohydr. Res.* **2005**, *340* (7), 1335–1342. https://doi.org/10.1016/j.carres.2005.03.004.

(5) Redgwell, R. J.; Melton, L. D.; Brasch, D. J. Changes to Pectic and Hemicellulosic Polysaccharides of Kiwifruit During Ripening. In *II International Symposium on Kiwifruit*; International Society for Horticultural Science: Palmerston North, 1991; pp 627–634.

(6) Knee, M. Polysaccharide Changes in Cell Walls of Ripening Apples. *Phytochemistry* **1973**, *12* (7), 1543–1549. https://doi.org/10.1016/0031-9422(73)80365-6.

(7) Guan, J.; Li, S. P. Discrimination of Polysaccharides from Traditional Chinese Medicines Using Saccharide Mapping-Enzymatic Digestion Followed by Chromatographic Analysis. *J. Pharm. Biomed. Anal.* **2010**, *51* (3), 590–598. https://doi.org/10.1016/j.jpba.2009.09.026.

(8) Jing, P.; Zhao, S. J.; Lu, M. M.; Cai, Z.; Pang, J.; Song, L. H. Multiple-Fingerprint Analysis for Investigating Quality Control of Flammulina Velutipes Fruiting Body Polysaccharides. *J. Agric. Food Chem.* **2014**, *62* (50), 12128–12133. https://doi.org/10.1021/jf504349r.

(9) Wang, Y.; Xian, J.; Xi, X.; Wei, X. Multi-Fingerprint and Quality Control Analysis of Tea Polysaccharides. *Carbohydr. Polym.* **2013**, *92* (1), 583–590. https://doi.org/10.1016/j.carbpol.2012.09.004.

(10) 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*, 4418–4424. https://doi.org/10.1021/acs.jafc.9b00720.

(11) Nie, S. P.; Wang, C.; Cui, S. W.; Wang, Q.; Xie, M. Y.; Phillips, G. O. A Further Amendment to the Classical Core Structure of Gum Arabic (Acacia Senegal). *Food Hydrocoll.* **2013**, *31* (1), 42–48. https://doi.org/10.1016/j.foodhyd.2012.09.014.

(12) Dourado, F.; Cardoso, S. M.; Silva, A. M. S.; Gama, F. M.; Coimbra, M. A. NMR Structural Elucidation of the Arabinan from Prunus Dulcis Immunobiological Active Pectic Polysaccharides. *Carbohydr. Polym.* **2006**, *66* (1), 27–33.

https://doi.org/10.1016/j.carbpol.2006.02.020.

(13) Merkx, D. W. H.; Westphal, Y.; van Velzen, E. J. J.; Thakoer, K. V.; de Roo, N.; van Duynhoven, J. P. M. Quantification of Food Polysaccharide Mixtures by 1H NMR. *Carbohydr. Polym.* **2018**, *179*, 379–385. https://doi.org/10.1016/j.carbpol.2017.09.074.

(14) Rocha, M.; Delgadillo, I.; Marcela, C.; Jana, C. Use of FT-IR Spectroscopy as a Tool for the Analysis of Polysaccharide Food Additives. **2003**, *51*, 383–389.

(15) Coimbra, M. A.; Gonçalves, F.; Barros, A. S.; Delgadillo, I. Fourier Transform Infrared Spectroscopy and Chemometric Analysis of White Wine Polysaccharide Extracts. *J. Agric. Food Chem.* **2002**, *50* (12), 3405–3411. https://doi.org/10.1021/jf020074p.

(16) Lerouxel, O.; Choo, T. S.; Se, M.; Lerouge, P.; Pauly, M. Rapid Structural Phenotyping of Plant Cell Wall Mutants by Enzymatic Oligosaccharide Fingerprinting. *Plant Physiol.* **2002**, *130*, 1754–1763. https://doi.org/10.1104/pp.011965.Progress.

(17) Obel, N.; Erben, V.; Schwarz, T.; Kühnel, S.; Fodor, A.; Pauly, M. Microanalysis of Plant Cell Wall Polysaccharides. *Mol. Plant* **2009**, *2* (5), 922–932. https://doi.org/10.1093/mp/ssp046.

(18) Doco, T.; O'Neill, M. A.; 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. *Carbohydr. Polym.* **2001**, *46* (3), 249–259. https://doi.org/10.1016/S0144-8617(00)00328-3.

(19) Guadalupe, Z.; Martínez-Pinilla, O.; Garrido, Á.; Carrillo, J. D.; Ayestarán, B. Quantitative Determination of Wine Polysaccharides by Gas Chromatography-Mass Spectrometry (GC-MS) and Size Exclusion Chromatography (SEC). *Food Chem.* **2012**, *131* (1), 367–374. https://doi.org/10.1016/j.foodchem.2011.08.049.

(20) Xia, Y. G.; Wang, T. L.; Sun, H. M.; Liang, J.; Kuang, H. X. Gas Chromatography–Mass Spectrometry-Based Trimethylsilyl-Alditol Derivatives for Quantitation and Fingerprint Analysis of Anemarrhena Asphodeloides Bunge Polysaccharides. *Carbohydr. Polym.* **2018**, *198* (February), 155–163. https://doi.org/10.1016/j.carbpol.2018.06.066.

(21) 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. https://doi.org/10.1016/j.ijms.2018.12.009.

(22) 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), 1–12. https://doi.org/10.1038/s41467-020-17778-1.

(23) Nishinari, K.; Takemasa, M.; Zhang, H.; Takahashi, R. Storage Plant Polysaccharides: Xyloglucans, Galactomannans, Glucomannans. In *Comprehensive glycoscience: from chemistry to systems biology*; Kamerling, H., Ed.; Elsevier, 2007; pp 613–652.

(24) Fischbacher, A.; von Sonntag, C.; Schmidt, T. C. Hydroxyl Radical Yields in the Fenton Process under Various PH, Ligand Concentrations and Hydrogen Peroxide/Fe(II) Ratios.

Chemosphere **2017**, *182*, 738–744. https://doi.org/10.1016/j.chemosphere.2017.05.039.

(25) Chan, K. H.; Chu, W. The Dose and Ratio Effects of Fe(II) and H_2O_2 in Fenton's Process on the Removal of Atrazine. *Environ. Technol.* **2003**, *24* (6), 703–710. https://doi.org/10.1080/09593330309385606.

(26) Barb, W. G.; Baxendale, J. H.; George, P.; Hargrave, K. R. Reactions of Ferrous and Ferric Ions with Hydrogen Peroxide Part I: The Ferrous Ion Reaction. *Trans. Faraday Soc.* **1951**, *47*, 462–500.

(27) Weiss, J.; Humphrey, C. W. Reaction between Hydrogen Peroxide and Iron Salts. *Nature* **1949**, *163* (4148), 691.

(28) 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.; et al. A Strategy for Annotating the Human Milk Glycome. *J. Agric. Food Chem.* **2006**, *54* (20), 7471–7480. https://doi.org/10.1021/jf0615810.

(29) West, C.; Elfakir, C.; Lafosse, M. Porous Graphitic Carbon: A Versatile Stationary Phase for Liquid Chromatography. *J. Chromatogr. A* **2010**, *1217* (19), 3201–3216. https://doi.org/10.1016/j.chroma.2009.09.052.

(30) Phillips, T. G. Changes in the Composition of Squash During Storage. *Plant Physiol.* **1946**, *21* (4), 533–541.

(31) Xu, S.; Yang, L.; Tian, R.; Wang, Z.; Liu, Z.; Xie, P.; Feng, Q. Species Differentiation and Quality Assessment of Radix Paeoniae Rubra (Chi-Shao) by Means of High-Performance Liquid Chromatographic Fingerprint. *J. Chromatogr. A* **2009**, *1216* (11), 2163–2168. https://doi.org/10.1016/j.chroma.2008.04.064.

(32) Sun, X.; Wang, H.; Han, X.; Chen, S.; Zhu, S.; Dai, J. Fingerprint Analysis of Polysaccharides from Different Ganoderma by HPLC Combined with Chemometrics Methods. *Carbohydr. Polym.* **2014**, *114*, 432–439. https://doi.org/10.1016/j.carbpol.2014.08.048.

(33) Khuwijitjaru, P.; Pokpong, A.; Klinchongkon, K.; Adachi, S. Production of Oligosaccharides from Coconut Meal by Subcritical Water Treatment. *Int. J. Food Sci. Technol.* **2014**, *49* (8), 1946–1952. https://doi.org/10.1111/ijfs.12524.

(34) Gwirtz, J. A.; Garcia-Casal, M. N. Processing Maize Flour and Corn Meal Food Products. *Ann. N. Y. Acad. Sci.* **2014**, *1312* (1), 66–75. https://doi.org/10.1111/nyas.12299.

(35) Zhu, K.; Zhang, Y.; Nie, S.; Xu, F.; He, S.; Gong, D.; Wu, G.; Tan, L. Physicochemical Properties and in Vitro Antioxidant Activities of Polysaccharide from Artocarpus Heterophyllus Lam. Pulp. *Carbohydr. Polym.* **2017**, *155*, 354–361. https://doi.org/10.1016/j.carbpol.2016.08.074.

(36) Tan, Y. F.; Li, H. L.; Lai, W. Y.; Zhang, J. Q. Crude Dietary Polysaccharide Fraction Isolated from Jackfruit Enhances Immune System Activity in Mice. *J. Med. Food* **2013**, *16* (7), 663–668. https://doi.org/10.1089/jmf.2012.2565.

(37) Jiménez-Escrig, A.; Rincón, M.; Pulido, R.; Saura-Calixto, F. Guava Fruit (Psidium

Guajava L.) as a New Source of Antioxidant Dietary Fiber. *J. Agric. Food Chem.* **2001**, *49* (11), 5489–5493. https://doi.org/10.1021/jf010147p.

(38) Zhang, Z.; Kong, F.; Ni, H.; Mo, Z.; Wan, J. B.; Hua, D.; Yan, C. Structural Characterization, α-Glucosidase Inhibitory and DPPH Scavenging Activities of Polysaccharides from Guava. *Carbohydr. Polym.* **2016**, *144*, 106–114. https://doi.org/10.1016/j.carbpol.2016.02.030.

(39) Myoda, T.; Matsuda, Y.; Suzuki, T.; Nakagawa, T.; Nagai, T.; Nagashima, T. Identification of Soluble Proteins and Interaction with Mannan in Mucilage of Dioscorea Opposita Thunb. (Chinese Yam Tuber). *Food Sci. Technol. Res.* **2006**, *12* (4), 299–302. https://doi.org/10.3136/fstr.12.299.

(40) Fu, Y. C.; Chen, S.; Lai, Y. J. Centrifugation and Foam Fractionation Effect on Mucilage Recovery from Dioscorea (Yam) Tuber. *J. Food Sci.* **2004**, *69* (9), 471–477.

(41) Yang, W.; Wang, Y.; Li, X.; Yu, P. Purification and Structural Characterization of Chinese Yam Polysaccharide and Its Activities. *Carbohydr. Polym.* **2015**, *117*, 1021–1027. https://doi.org/10.1016/j.carbpol.2014.09.082.

(42) Vollendorf, N. W.; Marlett, J. A. Comparison of Two Methods of Fiber Analysis of 58 Foods. *Journal of Food Composition and Analysis*. 1993, pp 203–214. https://doi.org/10.1006/jfca.1993.1023.

(43) Monono, E. M.; Haagenson, D. M.; Pryor, S. W. Developing and Evaluating NIR Calibration Models for Multi-Species Herbaceous Perennials. *Ind. Biotechnol.* **2012**, *8* (5), 285–292. https://doi.org/10.1089/ind.2012.0018.

(44) Monono, E. M.; Nyren, P. E.; Berti, M. T.; Pryor, S. W. Variability in Biomass Yield, Chemical Composition, and Ethanol Potential of Individual and Mixed Herbaceous Biomass Species Grown in North Dakota. *Ind. Crops Prod.* **2013**, *41* (1), 331–339. https://doi.org/10.1016/j.indcrop.2012.04.051.

(45) Zheng, Y.; Pan, Z.; Zhang, R.; Wang, D.; Labavitch, J.; Jenkins, B. M. Dilute Acid Pretreatment and Enzymatic Hydrolysis of Saline Biomass for Sugar Production. In *2006 ASABE Annual International Meeting*; American Society of Agricultural and Biological Engineers: Portland, OR, 2006.

(46) Wang, L. Z.; White, P. J. Structure and Properties of Amylose, Amylopectin, and Intermediate Materials of Oat Starches. *Cereal Chemistry*. 1994, pp 263–268.

(47) Pronyk, C.; Mazza, G. Fractionation of Triticale, Wheat, Barley, Oats, Canola, and Mustard Straws for the Production of Carbohydrates and Lignins. *Bioresour. Technol.* **2012**, *106*, 117–124. https://doi.org/10.1016/j.biortech.2011.11.071.

(48) Wood, P. J. Cereal β-Glucans in Diet and Health. *J. Cereal Sci.* **2007**, *46* (3), 230–238. https://doi.org/10.1016/j.jcs.2007.06.012.

(49) Butt, M. S.; Tahir-Nadeem, M.; Khan, M. K. I.; Shabir, R.; Butt, M. S. Oat: Unique among the Cereals. *Eur. J. Nutr.* **2008**, *47* (2), 68–79. https://doi.org/10.1007/s00394-008-0698-7.

(50) Varo, P.; Laine, R.; Veijalainen, K.; Espo, A.; Wetterhoff, A.; Koivistoinen, P. Dietary Fibre and Available Carbohydrates in Finnish Vegetables and Fruits. *J. Agric. Sci. Finl.* **1984**, *56*, 49–59.

(51) Ballesteros, L. F.; Cerqueira, M. A.; Teixeira, J. A.; Mussatto, S. I. Characterization of Polysaccharides Extracted from Spent Coffee Grounds by Alkali Pretreatment. *Carbohydr. Polym.* **2015**, *127*, 347–354. https://doi.org/10.1016/j.carbpol.2015.03.047.

(52) Pettolino, F. A.; Walsh, C.; Fincher, G. B.; Bacic, A. Determining the Polysaccharide Composition of Plant Cell Walls. *Nat. Protoc.* **2012**, *7* (9), 1590–1607. https://doi.org/10.1038/nprot.2012.081.

Chapter 3. Quantitative bottom-up glycomic analysis of polysaccharides in food matrices using liquid chromatography – tandem mass spectrometry

ABSTRACT

Carbohydrates are the most abundant biomolecules in nature and, specifically, polysaccharides are present in almost all plants and fungi. Due to their compositional diversity, polysaccharide analysis remains challenging. Compared to other biomolecules, high-throughput analysis for carbohydrates has yet to be developed. To address this gap in analytical science, we have developed a multiplexed, high-throughput, and quantitative approach for polysaccharide analysis in foods. Specifically, polysaccharides were depolymerized using a non-enzymatic chemical-digestion process followed by oligosaccharide fingerprinting using high performance liquid chromatography-quadrupole time-of-flight mass spectrometry (HPLC-QTOF-MS). Both label-free relative quantitation and absolute quantitation were done based on the abundances of oligosaccharides produced. Method validation included evaluating recovery for a range of polysaccharide standards and a breakfast cereal standard. Nine polysaccharides (starch, cellulose, β-glucan, mannan, galactan, arabinan, xylan, xyloglucan, chitin) were successfully quantitated with sufficient accuracy (5-25% bias) and high reproducibility (2-15% CV). Additionally, the method was used to identify and quantitate polysaccharides from a diverse sample set of food samples. Absolute concentrations of nine polysaccharides from apple and onions were obtained using an external calibration curve, where varietal differences were observed in some of the samples. The methodology developed in this study will provide complementary polysaccharide-level information to deepen our understanding of the interactions of dietary polysaccharides, gut microbial community, and human health.

INTRODUCTION

Carbohydrates are the most abundant class of biomolecules in nature[1], however, their analysis remains challenging. Polysaccharides in particular remain difficult to analyze because of their structural and compositional diversity. Food carbohydrates play an important role in human health, both directly (e.g. absorbed free sugars and products of gastrointestinal hydrolysis of starch) and indirectly from the impact of non-digestible components ("dietary fiber") on nutrient absorption and on the gut microbiome.[2] More recently, the effect of undigested polysaccharides (and oligosaccharides) in shaping and modulating the community of microbes in the human gut, and the effect on human health has been recognized and is a subject of widespread research efforts.[3,4] While endogenous human carbohydrate-active enzymes (CAZymes) are limited in function, gut microbes have a vast array of CAZymes that can potentially degrade polysaccharides and ferment them into secondary metabolites.[5] Different polysaccharide compositions and structures affect the gut microbiota in various ways owing to the taxonomical and functional diversity of these microbes.[6] Overall, changes in the gut microbiome induced by exposure to various polysaccharides can in turn induce metabolic and physiological changes in their host.[7,8] Detailed characterization of the food carbohydrates, specifically their chemical structures, is indispensable in establishing the relationship between food and health but analytical methods for comprehensive polysaccharide characterization are lacking.[9]

Starch and non-starch polysaccharides in foods are typically measured indirectly by enzymatic-gravimetric methods (e.g. AOAC 991.43, AOAC 2011.25, AOAC 2017.16) to obtain food composition data. When specific polysaccharides are characterized, they are typically extracted from biological sources first, and then fractionated by different buffers based on solubility. These fractions are then subjected separately to monosaccharide and linkage analyses

and the polysaccharide structures are inferred.[10,11] NMR techniques can be performed to confirm the primary structures of the purified polysaccharides.[12,13] Although this approach can provide an in-depth structural analysis, it is impractical for large-scale analysis of many foods and food products. NMR has also been recently used for absolute quantitation of some common polysaccharides. However, this specific method required the molar stoichiometry of monosaccharides in the mixture.[14] Other methodology has involved the use of CAZymes to deduce polysaccharide structure, where oligosaccharide products from selective enzymatic digestion are in turn characterized using chromatography and/or mass spectrometry (MS).[10,15] However, each enzyme reaction often requires optimization, rendering the method highly laborious with very low throughput. Monoclonal antibodies have also been developed and used to detect specific polysaccharides in plant tissues. This assay is typically performed in a microarray format where the extracted polysaccharide fractions are immobilized on multiple substrates to allow antibody binding.[16,17] While the method can have high throughput, limitations include the cost and availability of the antibodies, and extensive matrix effects of native samples.

To address the lack of a widely applicable and high-throughput method for quantitative polysaccharide analysis in foods, we have developed a method using a bottom-up glycomics approach (**Figure 3.1)**. Polysaccharide identification was based on the generation of characteristic oligosaccharides that were produced using Fenton chemistry in a reaction called "**F**enton's **I**nitiation **T**owards **D**efined **O**ligosaccharide **G**roups" (FITDOG).[18,19] The oligosaccharides were used as fingerprinting features to identity and quantitate polysaccharides based on chromatographic and tandem MS (MS/MS) analysis, where MS/MS provides compositional analysis of the oligosaccharides and chromatographic retention times facilitate further identification, with peak areas used for quantitation of the parent polysaccharides. The

methodology presented here significantly improves on our previously published workflow. The ability to simultaneously measure absolute concentration of nine polysaccharides in a single method is unprecedented. This approach was validated using standards and was applied to a variety of food types to identify and quantitate polysaccharides, in terms of both relative and absolute concentrations.

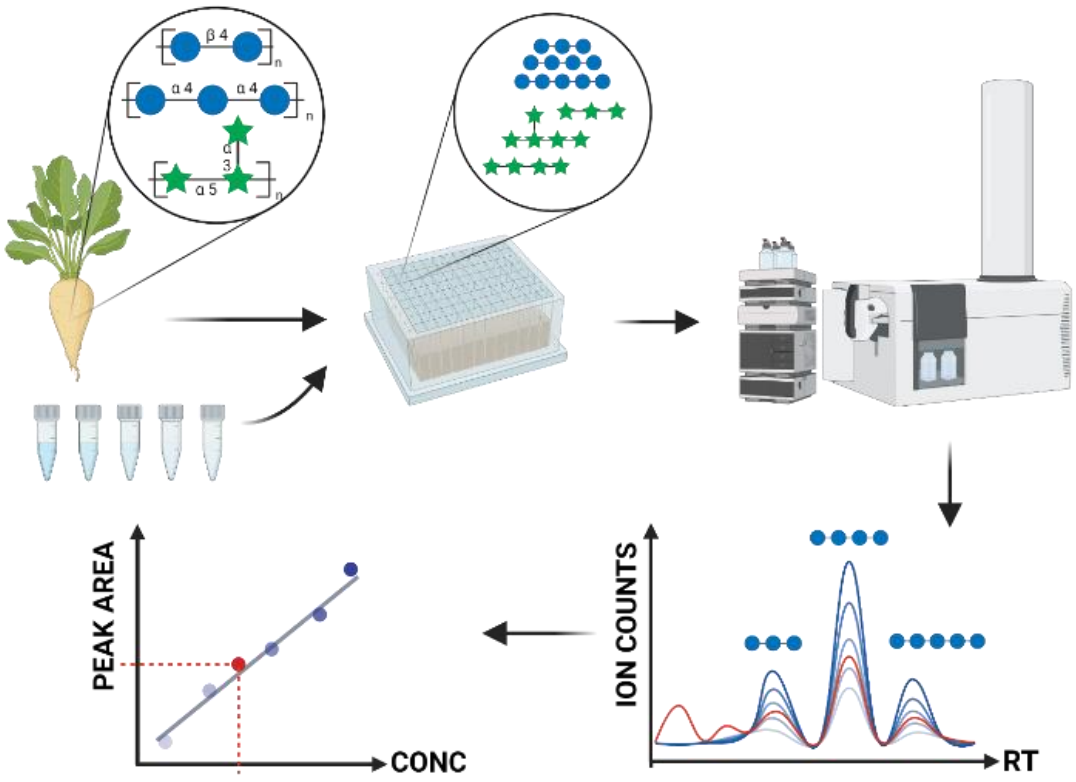


Figure 3.5. Overview of the analytical method for the identification and quantitation of polysaccharides using FITDOG and HPLC-QTOF profiling of the resulting oligosaccharides. The peak areas and the use of external calibration curves provided absolute quantitation.

Methods

Materials and reagents

Sodium acetate ($Na(CH_3CO_2)$), hydrogen peroxide (30% H_2O_2), sodium hydroxide (NaOH), iron(III) sulfate pentahydrate ($Fe_2(SO_4)_3 \cdot 5H_2O$), chitin (shrimp shells, BioReagent grade), and starch (corn, analytical grade) were purchased from Sigma-Aldrich (St. Louis, MO, USA). Cellulose (microcrystalline powder, extra pure, average particle size 90 μm) was purchased from ACROS Organics. Linear arabinan (sugar beet pulp, purity > 95%), mannan (ivory nut seeds, purity > 98%), galactan (potato fiber, purity > 85%), xylan (beechwood, purity > 95%), xyloglucan (tamarind seeds, purity > 95%), and β-glucan (barley flour, purity ~95%) were purchased from Megazyme (Bray, Ireland). Sodium borohydride ($NaBH_4$, powder, > 99%), trifluoroacetic acid (TFA, LC-MS grade) and formic acid (FA, LC-MS grade) were purchased from Fisher Scientific (Belgium, UK). Acetonitrile (ACN, HPLC grade) was purchased from Honeywell (Muskegon, MI). Nanopure water (18.2 MΩ-cm) was used for all experiments. Various fruits, vegetables, and herbs were prepared for method testing and were purchased from local grocery stores in Davis, CA, USA. The products were selected to represent ones containing a range of polysaccharide (and other saccharides) types and levels. Apples (Red Delicious, Honeycrisp, Granny Smith, Gala, Fuji) and onions (red, yellow, white) were procured and analyzed for the USDA Food DataCentral Foundation Foods database (https://fdc.nal.usda.gov) from different retail stores in the Beltsville, MD and Blackburg, VA areas in 2020. Eight samples of each food/variety were obtained, with each sample being approximately 1-1.5 kg total. Apples were analyzed with skin but without stem and core, and onions were analyzed without skin. Preparation of homogenates in liquid nitrogen and storage of the prepared subsamples was as described previously.[21] Solid-phase extraction cartridges (C18 and PGC) in 96-well plate format

were purchased from Glygen Corporation (Columbia, MD, USA). Reaction plates in deep 96-well format (Nunc™ 96-Well Polypropylene DeepWell™ Storage Plates) were purchased from Thermo Scientific (Waltham, MA, USA).

Food sample preparation

At least 5.0 g of fresh food sample was weighed in a 50-mL screwcap tube, frozen, and freeze-dried (BenchTop Pro, SP Scientific, Warminster, PA, USA) for at least 48 hr. Moisture content was determined from the fresh weight and the dried weight after freeze drying (residual moisture was not measured and was assumed to be zero). Dried food samples were then ground into powder using 3.2-mm stainless steel beads and homogenized using the Bead Ruptor Elite Bead Mill Homogenizer (Omni International, Kennesaw, GA, USA). Each dried and ground food sample was weighed (25 mg), suspended in 1.00 mL 80% ethanol in a 1.5-mL screwcap tube, homogenized using the bead mill homogenizer, and centrifuged at 15,000 rcf for 15 min. Supernatant was discarded and the pellet residue was further washed twice using 1.00 mL 80% ethanol using the same conditions as the first wash. Ethanol-washed pellet was dried in a centrifugal vacuum evaporator. Dried pellets were then suspended in 1.00 mL water, homogenized with 0.9-mm stainless steel beads, heated at 100 ^{o}C for 1 hr using an incubator oven (OF-01E, Jeio Tech, Daejeon, Republic of Korea) without shaking, and then homogenized again with the bead mill homogenizer. A 100-µL aliquot was plated into 96 deep-well plate for the FITDOG reaction. From this step onwards, all steps were carried out in 96-well plate format, allowing for a rapid throughput and scalable method.

Depolymerization reaction using Fenton's reagent

We have previously optimized this reaction to yield reproducible and diverse oligosaccharides.[19] The reaction mixture consisted of 95% (v/v) 44 mM sodium acetate buffer (adjusted to pH 5.20 with glacial acetic acid), 5% (v/v) of 30% (v/v) H_2O_2, and 65 nM $Fe_2(SO_4)_3 \cdot 5H_2O$. To each of the well, an aliquot of 100 μL of sample or standard mixture was transferred, then 900 μL of the reaction mixture was added and allowed to react for 1 hr at 100 ^oC using incubator oven without shaking. The reaction was quenched by adding 500 μL of freshly prepared 2 M NaOH, followed by glacial acetic acid (61 μL) for neutralization. The resulting oligosaccharides were then reduced by incubation with an equal volume of 1 M $NaBH_4$ for 1 hr at 65 ^oC, followed by isolation and clean-up using sequential solid-phase extractions (SPE) in a 96-well plate format. Samples were cleaned up first with C18 SPE, then porous graphitized carbon (PGC) SPE. The recovered and cleaned-up oligosaccharides were completely dried by centrifugal vacuum evaporator and stored at -20 ^oC until analysis.

Calibration standards preparation

Multiplexed quantitation of polysaccharides was enabled by pooling several polysaccharide standards together. Three mixtures of polysaccharides were prepared by the following scheme: mixture 1 contains arabinan, galactan, β-glucan; mixture 2 contains xylan, mannan, xyloglucan, chitin; mixture 3 contains starch and cellulose. Each polysaccharide was weighed (~ 10 mg) in a 7-mL polypropylene vial and each mixture was suspended in 5-mL water. Pooled mixtures were homogenized with 0.9-mm stainless steel beads in a bead mill homogenizer, heated at 100 ^oC for 1 hr, and then homogenized again with the bead mill

homogenizer. Calibration curve standards were prepared by serial dilutions of the pooled mixtures. From these, 100-µL aliquots were transferred to 96 deep-well reaction plate, together with the processed food samples for the FITDOG reaction. Mixtures used for validation were made from the same pooled mixtures but diluted at different concentrations. To assess the reproducibility of each step of the workflow, a pool of five polysaccharides (starch, cellulose, arabinan, xylan, chitin at ~2 mg/mL of each) was prepared similarly in water, homogenized, heated at 100 ^{o}C for 1 hr, and then homogenized again with the bead mill homogenizer. This pooled mixture was plated in 9 wells to assess the overall method variability. In each subsequent steps of the procedure ($NaBH_4$ reduction, clean-up, instrument injection), several aliquots from the previous steps were pooled and were used as new replicates for the next steps.

Solid phase extraction (SPE) clean-up

Reduced oligosaccharides were cleaned up with C18 SPE first, then porous graphitized carbon (PGC) SPE. Each step of the SPE protocol was carried out with a centrifuge (Centrifuge 5810-R, Eppendorf, Hamburg, Germany) with deep-well plate rotor (1,200 rcf for 1 min). For the C18 SPE, cartridges were washed first with ACN (200 µL, 2×), then water (200 µL, 2×). The sample solution was then loaded (400 µL, 2×) and flow-through was collected. For the PGC SPE, cartridges were primed with water (400 µL, 1×), then 80% ACN, 0.1% (v/v) TFA (400 µL, 1×) and finally water (400 µL, 1×). The eluent from C18 SPE was loaded (400 µL, 2×) and washed with water (400 µL, 6×). The oligosaccharides were then eluted with 400 µL 40% ACN with 0.05% (v/v) TFA in a 0.8-mL 96-well collection plate (Abgene™ 96 Well 0.8mL Polypropylene Deepwell Storage Plate, Thermo Scientific). The recovered eluent was completely dried by centrifugal vacuum evaporator and stored at -20 ^{o}C until analysis.

Liquid chromatography – mass spectrometry

Samples were reconstituted in 100 µL water (18 MΩ) prior to analytical separation, which was carried out using an Agilent 1260 Infinity II HPLC (Agilent Technologies, Santa Clara, CA, USA). Chromatographic separation was performed on a 150 mm × 1 mm Hypercarb column from Thermo Scientific (5-µm particle size). The column compartment was set at 40 ^{o}C. A binary gradient was employed and consisted of solvent A: (3 % (v/v) ACN, 0.1 % FA in water) and solvent B: (90 % ACN, 0.1 % FA in water). A 45-min gradient with a flow rate of 0.132 mL/min was used: 3-25 % B, 0-15 min; 25-25 % B, 15-18 min; 25-99 % B, 18-30 min; 99-99 %B, 30-32 min; 99-3 % B, 32-34 min; 3-3 % B, 34-45 min.

HPLC was coupled to Agilent 6530 Accurate-Mass Q-TOF mass spectrometer (Agilent Technologies, Santa Clara, CA, USA). The MS detector was run in the positive mode with the following electrospray source parameters: drying gas temperature = 150 ^{o}C. drying gas flow rate = 11 L/min, fragmentor = 175 V, skimmer = 60 V, octupole 1 RF = 750 V. Acquisition mode was set to data-dependent mode, where top 5 most abundant precursor ions were selected for fragmentation. Dynamic exclusion was enabled for 30 s. The acquisition rate was set to 0.63 spectra/s. For tandem MS fragmentation, a linear function for collision energy (CE), where CE = 1.45*(*m/z*)-3.5, was employed.

Data analysis

The FITDOG reaction was used to generate distinct oligosaccharides from parent polysaccharides. Individual polysaccharide standards were prepared at ~2.00 mg/mL in water, homogenized using the bead mill homogenizer, and then heated at 100 ^{o}C for 1 hr.

Polysaccharides were then reacted using the FITDOG process and the resulting oligosaccharides were used to construct the fingerprint library (**Table 3.1**). Tandem mass fragmentation spectra were used to manually assign monosaccharide class compositions to each oligosaccharide peak. Chromatographic retention times and accurate masses were used to match sample oligosaccharides to the library.

Table 3.1. Oligosaccharide (reduced) fingerprinting library for identification and quantitation of polysaccharides using the FITDOG workflow. Retention time values were based on the LC conditions described in the Methods section. Oligosaccharide compositions were deduced from tandem mass spectra fragmentation pattern.

Polysaccharide	Oligo-saccharide	Formula (reduced)	RT(min)	Used as quantifier?
Starch	Hex3	C18 H34 O16	3.40	No
Starch	Hex4	C24 H44 O21	11.75	Yes
Starch	Hex5	C30 H54 O26	13.70	Yes
Starch	Hex6	C36 H64 O31	14.59	Yes
Starch	Hex7	C42 H74 O36	15.09	No
Starch	Hex8	C48 H84 O41	15.59	No
Starch	Hex9	C54 H94 O46	16.26	No
Starch	Hex10	C60 H104 O51	17.26	No
Starch	Hex11	C66 H114 O56	18.49	No
Starch	Hex12	C72 H124 O61	19.49	No
Starch	Hex13	C78 H134 O66	20.10	No
Starch	Hex14	C84 H144 O71	20.55	No
Starch	Hex15	C90 H154 O76	20.88	No
Starch	Hex16	C96 H164 O81	21.27	No
Starch	Hex17	C102 H174 O86	22.11	No
Starch	Hex18	C108 H184 O91	23.28	No
Starch	Hex19	C114 H194 O96	24.40	No
Starch	Hex20	C120 H204 O101	24.77	No
Starch	Hex21	C126 H214 O106	24.89	No
Starch	Hex22	C132 H224 O111	25.03	No
Cellulose	Hex3	C18 H34 O16	14.75	Yes
Cellulose	Hex4	C24 H44 O21	20.21	Yes
Cellulose	Hex5	C30 H54 O26	25.73	Yes
Mannan	Hex3	C18 H34 O16	1.67	No
Mannan	Hex4	C24 H44 O21	3.40	Yes
Mannan	Hex5	C30 H54 O26	9.97	Yes
Mannan	Hex6	C36 H64 O31	12.93	Yes

Mannan	Hex7	C42 H74 O36	13.93	No
Mannan	Hex8	C48 H84 O41	14.77	No
Mannan	Hex9	C54 H94 O46	15.43	No
Mannan	Hex10	C60 H104 O51	15.94	No
Mannan	Hex11	C66 H114 O56	16.49	No
Mannan	Hex12	C72 H124 O61	16.83	No
Mannan	Hex13	C78 H134 O66	17.27	No
Mannan	Hex14	C84 H144 O71	17.72	No
b-Glucan	Hex3	C18 H34 O16	12.59	No
b-Glucan	Hex3	C18 H34 O16	14.75	No
b-Glucan	Hex3	C18 H34 O16	16.27	No
b-Glucan	Hex4	C24 H44 O21	18.33	No
b-Glucan	Hex4	C24 H44 O21	20.39	No
b-Glucan	Hex4	C24 H44 O21	21.23	Yes
b-Glucan	Hex5	C30 H54 O26	23.46	No
b-Glucan	Hex5	C30 H54 O26	24.57	No
b-Glucan	Hex5	C30 H54 O26	25.74	No
b-Glucan	Hex5	C30 H54 O26	26.52	Yes
b-Glucan	Hex5	C30 H54 O26	27.30	Yes
Xylan	Pnt3	C15 H28 O13	10.87	Yes
Xylan	Pnt4	C20 H36 O17	16.21	Yes
Xylan	Pnt5	C25 H44 O21	19.45	Yes
Xylan	Pnt6	C30 H52 O25	22.40	No
Xylan	Pnt7	C35 H60 O29	25.96	No
Xylan	Pnt8	C40 H68 O33	28.08	No
Chitin	HexNAc3	C16 H30 O11 N2	10.60	Yes
Chitin	HexNAc4	C24 H43 O16 N3	14.80	No
Chitin	HexNAc5	C32 H56 O21 N4	16.30	Yes
Chitin	HexNAc6	C40 H69 O26 N5	18.20	Yes
Chitin	HexNAc7	C48 H82 O31 N6	19.10	No
Chitin	HexNAc8	C56 H95 O36 N7	20.40	No
Chitin	HexNAc9	C64 H108 O41 N8	21.12	No
Arabinan, linear	Pnt3	C15 H28 O13	5.18	Yes
Arabinan, linear	Pnt4	C20 H36 O17	13.42	Yes
Arabinan, linear	Pnt5	C25 H44 O21	15.92	Yes
Arabinan, linear	Pnt6	C30 H52 O25	17.76	No
Arabinan, linear	Pnt7	C35 H60 O29	19.26	No
Arabinan, linear	Pnt8	C40 H68 O33	20.66	No
Arabinan, linear	Pnt9	C50 H84 O41	21.88	No
Galactan	Hex3	C12 H24 O11	2.28	Yes
Galactan	Hex4	C18 H34 O16	7.35	Yes
Galactan	Hex5	C24 H44 O21	12.20	Yes
Galactan	Hex6	C30 H54 O26	13.31	No

Galactan	Hex7	C36 H64 O31	13.92	No
Galactan	Hex8	C42 H74 O36	14.47	No
Galactan	Hex9	C48 H84 O41	15.37	No
Galactan	Hex10	C54 H94 O46	17.00	No
Galactan	Hex11	C60 H104 O51	18.40	No
Galactan	Hex12	C66 H114 O56	19.40	No
Galactan	Hex13	C72 H124 O61	20.00	No
Galactan	Hex14	C78 H134 O66	20.50	No
Xyloglucan	Hex2:Pnt1	C17 H32 O15	12.10	Yes
Xyloglucan	Hex2:Pnt2	C22 H40 O19	14.60	Yes
Xyloglucan	Hex3:Pnt1	C23 H42 O20	14.00	Yes
Xyloglucan	Hex3:Pnt1	C23 H42 O20	16.50	No
Xyloglucan	Hex3:Pnt1	C23 H42 O20	18.10	No
Xyloglucan	Hex3:Pnt2	C28 H50 O24	18.80	No
Arabinogalactan	Hex3	C18 H34 O16	1.73	No
Arabinogalactan	Hex2:Pnt1	C17 H32 O15	2.18	No
Arabinogalactan	Hex3	C18 H34 O16	2.3	No
Arabinogalactan	Hex4	C24 H44 O21	2.85	No
Arabinogalactan	Hex1:Pnt3	C21 H38 O18	3.21	No
Arabinogalactan	Hex4	C24 H44 O21	3.37	No
Arabinogalactan	Hex2:Pnt2	C22 H40 O19	3.82	No
Arabinogalactan	Pnt3	C15 H28 O13	4	No
Arabinogalactan	Hex4	C24 H44 O21	4.68	No
Arabinogalactan	Hex3	C18 H34 O16	4.83	No
Arabinogalactan	Hex5	C30 H54 O26	6.59	No
Arabinogalactan	Hex4	C24 H44 O21	6.95	No
Arabinogalactan	Hex5	C30 H54 O26	7.28	No
Arabinogalactan	Hex3:Pnt1	C23 H42 O20	7.28	No
Arabinogalactan	Hex2:Pnt1	C17 H32 O15	9.22	No
Arabinogalactan	Hex2:Pnt1	C17 H32 O15	11.21	No
Arabinogalactan	Hex6	C36 H64 O31	11.36	No
Arabinogalactan	Hex5	C30 H54 O26	11.5	No
Arabinogalactan	Hex2:Pnt1	C17 H32 O15	11.66	No
Arabinogalactan	Hex4	C24 H44 O21	12.14	No
Arabinogalactan	Hex3:Pnt1	C23 H42 O20	12.26	No
Arabinogalactan	Hex5	C30 H54 O26	12.8	No
Arabinogalactan	Hex4	C24 H44 O21	12.97	No
Arabinogalactan	Hex3:Pnt1	C23 H42 O20	13.31	No
Arabinogalactan	Hex2:Pnt2	C22 H40 O19	13.31	No
Arabinogalactan	Hex5	C30 H54 O26	13.48	No
Arabinogalactan	Hex6	C36 H64 O31	13.63	No
Arabinogalactan	Hex7	C42 H74 O36	14.3	No
Arabinogalactan	Hex6	C36 H64 O31	15	No

Arabinogalactan	Hex5:Pnt1	C35 H62 O30	15	No
Arabinogalactan	Hex6	C36 H64 O31	15.44	No
Arabinogalactan	Hex7	C42 H74 O36	16.25	No
Arabinogalactan	Hex6	C36 H64 O31	16.37	No
Arabinogalactan	Pnt4	C20 H36 O17	16.51	No
Arabinogalactan	Hex7	C42 H74 O36	16.69	No
Arabinogalactan	Hex8	C48 H84 O41	17	No
Arabinogalactan	Hex4:Pnt1	C29 H52 O25	17.19	No
Arabinogalactan	Hex7	C42 H74 O36	17.54	No
Arabinogalactan	Hex8	C48 H84 O41	18.1	No
Arabinogalactan	Hex8	C48 H84 O41	18.44	No
Arabinogalactan	Hex5	C30 H54 O26	18.81	No
Arabinoxylan	Pnt3	C15 H28 O13	2.04	No
Arabinoxylan	Pnt3	C15 H28 O13	2.67	No
Arabinoxylan	Pnt3	C15 H28 O13	4.23	No
Arabinoxylan	Pnt4	C20 H36 O17	4.74	No
Arabinoxylan	Pnt4	C20 H36 O17	7.48	No
Arabinoxylan	Pnt4	C20 H36 O17	8.23	No
Arabinoxylan	Pnt4	C20 H36 O17	10.76	No
Arabinoxylan	Pnt4	C20 H36 O17	12.07	No
Arabinoxylan	Pnt4	C20 H36 O17	12.38	No
Arabinoxylan	Pnt4	C20 H36 O17	13.15	No
Arabinoxylan	Pnt5	C25 H44 O21	13.41	No
Arabinoxylan	Pnt5	C25 H44 O21	13.69	No
Arabinoxylan	Pnt5	C25 H44 O21	14.25	No
Arabinoxylan	Pnt5	C25 H44 O21	14.55	No
Arabinoxylan	Pnt5	C25 H44 O21	14.91	No
Arabinoxylan	Pnt5	C25 H44 O21	15.59	No
Arabinoxylan	Pnt5	C25 H44 O21	16.11	No
Arabinoxylan	Pnt3	C15 H28 O13	16.47	No
Arabinoxylan	Pnt6	C30 H52 O25	16.6	No
Arabinoxylan	Pnt5	C25 H44 O21	17.05	No
Arabinoxylan	Pnt6	C30 H52 O25	17.3	No
Arabinoxylan	Pnt5	C25 H44 O21	18.03	No
Arabinoxylan	Pnt4	C20 H36 O17	18.45	No
Arabinoxylan	Pnt5	C25 H44 O21	18.53	No
Arabinoxylan	Pnt6	C30 H52 O25	18.77	No
Arabinoxylan	Pnt4	C20 H36 O17	19.13	No
Arabinoxylan	Pnt6	C30 H52 O25	19.65	No
Arabinoxylan	Pnt4	C20 H36 O17	19.75	No
Arabinoxylan	Pnt6	C30 H52 O25	20.24	No
Arabinoxylan	Pnt6	C30 H52 O25	21.16	No
Arabinoxylan	Pnt6	C30 H52 O25	21.39	No

Arabinoxylan	Pnt7	C35 H60 O29	21.67	No
Arabinoxylan	Pnt6	C30 H52 O25	21.94	No
Arabinoxylan	Pnt5	C25 H44 O21	22.06	No
Arabinoxylan	Pnt7	C35 H60 O29	22.59	No
Arabinoxylan	Pnt6	C30 H52 O25	22.71	No
Arabinoxylan	Pnt6	C30 H52 O25	23.14	No
Arabinoxylan	Pnt5	C25 H44 O21	24.41	No
Arabinoxylan	Pnt7	C35 H60 O29	24.55	No
Arabinoxylan	Pnt7	C35 H60 O29	25.37	No
Arabinoxylan	Pnt6	C30 H52 O25	26.06	No
Arabinoxylan	Pnt8	C40 H68 O33	27.62	No
Arabinoxylan	Pnt7	C35 H60 O29	28.46	No
Curdlan	Hex3	C18 H34 O16	13.7	No
Curdlan	Hex4	C24 H44 O21	20.72	No
Curdlan	Hex5	C30 H54 O26	27.96	No
Galactomannan	Hex3	C18 H34 O16	2.58	No
Galactomannan	Hex4	C24 H44 O21	3.88	No
Galactomannan	Hex4	C24 H44 O21	4.25	No
Galactomannan	Hex4	C24 H44 O21	4.67	No
Galactomannan	Hex4	C24 H44 O21	8.18	No
Galactomannan	Hex5	C30 H54 O26	9.99	No
Galactomannan	Hex5	C30 H54 O26	11.02	No
Galactomannan	Hex5	C30 H54 O26	11.34	No
Galactomannan	Hex4	C24 H44 O21	12.14	No
Galactomannan	Hex5	C30 H54 O26	12.26	No
Galactomannan	Hex6	C36 H64 O31	13.03	No
Galactomannan	Hex5	C30 H54 O26	13.15	No
Galactomannan	Hex6	C36 H64 O31	13.57	No
Galactomannan	Hex6	C36 H64 O31	13.79	No
Galactomannan	Hex6	C36 H64 O31	14.1	No
Galactomannan	Hex7	C42 H74 O36	14.63	No
Galactomannan	Hex6	C36 H64 O31	14.85	No
Galactomannan	Hex4	C24 H44 O21	15.21	No
Galactomannan	Hex7	C42 H74 O36	15.35	No
Galactomannan	Hex8	C48 H84 O41	15.9	No
Galactomannan	Hex6	C36 H64 O31	16.48	No
Galactomannan	Hex7	C42 H74 O36	16.63	No
Galactomannan	Hex8	C48 H84 O41	16.93	No
Glucomannan	Hex3	C18 H34 O16	1.92	No
Glucomannan	Hex4	C24 H44 O21	5.62	No
Glucomannan	Hex3	C18 H34 O16	6.55	No
Glucomannan	Hex3	C18 H34 O16	8.73	No
Glucomannan	Hex4	C24 H44 O21	10.62	No

Glucomannan	Hex5	C30 H54 O26	11.45	No
Glucomannan	Hex4	C24 H44 O21	11.6	No
Glucomannan	Hex3	C18 H34 O16	11.84	No
Glucomannan	Hex4	C24 H44 O21	12.27	No
Glucomannan	Hex4	C24 H44 O21	12.92	No
Glucomannan	Hex6	C36 H64 O31	12.93	No
Glucomannan	Hex5	C30 H54 O26	13.2	No
Glucomannan	Hex6	C36 H64 O31	13.56	No
Glucomannan	Hex3	C18 H34 O16	14.02	No
Glucomannan	Hex6	C36 H64 O31	14.31	No
Glucomannan	Hex7	C42 H74 O36	15.02	No
Glucomannan	Hex5	C30 H54 O26	15.26	No
Glucomannan	Hex6	C36 H64 O31	15.52	No
Glucomannan	Hex8	C48 H84 O41	15.82	No
Glucomannan	Hex5	C30 H54 O26	15.94	No
Glucomannan	Hex4	C24 H44 O21	16.23	No
Glucomannan	Hex6	C36 H64 O31	16.46	No
Glucomannan	Hex5	C30 H54 O26	16.62	No
Glucomannan	Hex4	C24 H44 O21	16.77	No
Glucomannan	Hex7	C42 H74 O36	17.02	No
Glucomannan	Hex5	C30 H54 O26	17.14	No
Glucomannan	Hex6	C36 H64 O31	17.3	No
Glucomannan	Hex5	C30 H54 O26	17.62	No
Glucomannan	Hex6	C36 H64 O31	17.74	No
Glucomannan	Hex5	C30 H54 O26	17.86	No
Glucomannan	Hex4	C24 H44 O21	18.01	No
Glucomannan	Hex5	C30 H54 O26	18.26	No
Glucomannan	Hex7	C42 H74 O36	18.37	No
Glucomannan	Hex6	C36 H64 O31	18.51	No
Glucomannan	Hex7	C42 H74 O36	18.79	No
Glucomannan	Hex5	C30 H54 O26	19.06	No
Glucomannan	Hex6	C36 H64 O31	19.14	No
Glucomannan	Hex7	C42 H74 O36	19.24	No
Glucomannan	Hex6	C36 H64 O31	19.4	No
Glucomannan	Hex7	C42 H74 O36	20.03	No
Glucomannan	Hex7	C42 H74 O36	20.31	No
Glucomannan	Hex8	C48 H84 O41	20.83	No
Glucomannan	Hex4	C24 H44 O21	21.11	No
Glucomannan	Hex7	C42 H74 O36	21.54	No
Glucomannan	Hex5	C30 H54 O26	21.85	No
Glucomannan	Hex7	C42 H74 O36	21.99	No
Glucomannan	Hex5	C30 H54 O26	22.27	No
Glucomannan	Hex6	C36 H64 O31	22.76	No

Glucomannan	Hex7	C42 H74 O36	23.22	No
Glucomannan	Hex7	C42 H74 O36	23.88	No
Glucomannan	Hex6	C36 H64 O31	24.72	No
Glucomannan	Hex5	C30 H54 O26	25.42	No
Glucomannan	Hex6	C36 H64 O31	25.77	No
Glucomannan	Hex8	C48 H84 O41	27	No
Glucomannan	Hex9	C54 H94 O46	27.68	No
Lichenan	Hex3	C18 H34 O16	2.04	No
Lichenan	Hex4	C24 H44 O21	2.88	No
Lichenan	Hex3	C18 H34 O16	3.06	No
Lichenan	Hex3	C18 H34 O16	3.34	No
Lichenan	Hex3	C18 H34 O16	5.85	No
Lichenan	Hex5	C30 H54 O26	6.14	No
Lichenan	Hex5	C30 H54 O26	7.59	No
Lichenan	Hex4	C24 H44 O21	9.07	No
Lichenan	Hex6	C36 H64 O31	10.1	No
Lichenan	Hex6	C36 H64 O31	10.78	No
Lichenan	Hex5	C30 H54 O26	11.08	No
Lichenan	Hex7	C42 H74 O36	12.13	No
Lichenan	Hex3	C18 H34 O16	12.59	No
Lichenan	Hex6	C36 H64 O31	12.89	No
Lichenan	Hex5	C30 H54 O26	13.42	No
Lichenan	Hex5	C30 H54 O26	14.29	No
Lichenan	Hex6	C36 H64 O31	14.75	No
Lichenan	Hex3	C18 H34 O16	15.02	No
Lichenan	Hex6	C36 H64 O31	15.34	No
Lichenan	Hex5	C30 H54 O26	16.53	No
Lichenan	Hex7	C42 H74 O36	16.62	No
Lichenan	Hex6	C36 H64 O31	17.23	No
Lichenan	Hex6	C36 H64 O31	17.62	No
Lichenan	Hex7	C42 H74 O36	18.1	No
Lichenan	Hex7	C42 H74 O36	18.52	No
Lichenan	Hex8	C48 H84 O41	18.96	No
Lichenan	Hex7	C42 H74 O36	20.39	No
Lichenan	Hex8	C48 H84 O41	20.99	No
Lichenan	Hex9	C54 H94 O46	21.45	No
Lichenan	Hex4	C24 H44 O21	22.22	No
Lichenan	Hex5	C30 H54 O26	24.88	No
Lichenan	Hex5	C30 H54 O26	25.82	No
Lichenan	Hex6	C36 H64 O31	27.31	No
Lichenan	Hex6	C36 H64 O31	31.06	No

For annotation of oligosaccharide peaks from food samples, an in-house script was used. Raw data was first converted to MGF (Mascot Generic Format) files to be parsed by GlycoNote, a Python script previously developed in our laboratory for automated glycan composition annotation from tandem MS spectra (https://github.com/MingqiLiu/GlycoNote). The script generates a combinatorial library of oligosaccharides from an input of possible monosaccharide class compositions. Tandem mass spectra from each sample were filtered based on precursor ion *m*/*z* values generated from the combinatorial library. Additional diagnostic ions could also be included to filter out non-oligosaccharide spectra. Tandem mass-spectral peaks were annotated based on commonly observed fragment ions of oligosaccharides (A_i, B_i, C_i, X_i, Y_i, Z_i) resulting from collision-induced dissociation (CID) fragmentation. Lastly, identification results are false-discovery rate-controlled by implementing a target-decoy strategy. The script outputs several files, including image files for each of the annotated tandem mass spectra (**Figure 3.2**), and a summary table which lists all the oligosaccharides identified (**Table 3.2**). GlycoNote is especially useful in large batch analysis. For this specific method, mass tolerances of 20 and 50 ppm were used for the precursor and fragment ions, respectively. The output list of compounds was filtered to >50% coverage based on intensity and monosaccharide sequence.

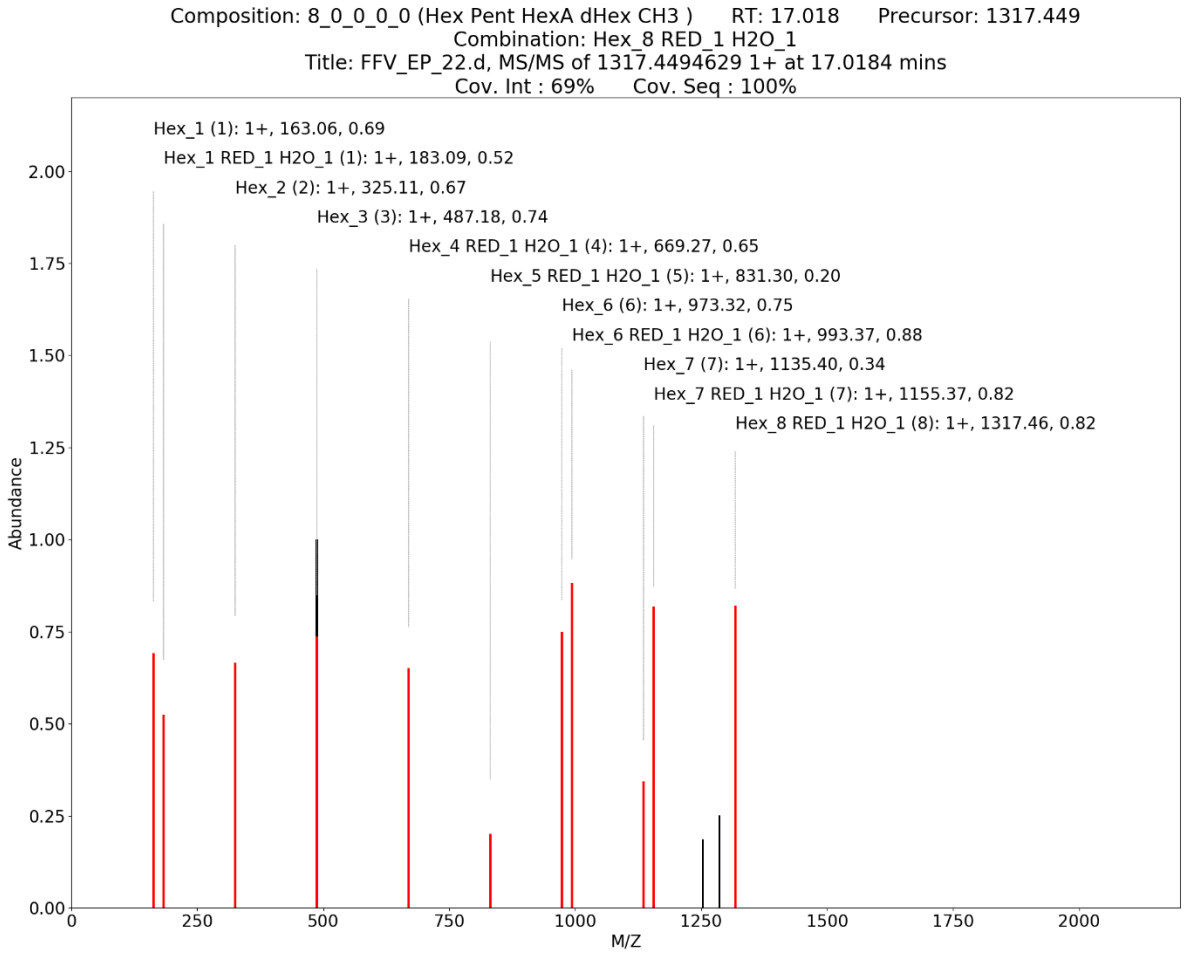


Figure 3.2. Example of a GlycoNote-annotated tandem mass spectrum of an Hex_8 oligosaccharide from the amylose standard.

Table 3.2. Example of annotation output of GlycoNote.

spectra	retention	precursor	glycan mass	PPM	Hex Pent HexA dHex CH3	description	charge	Cov. Int	Cov. Seq	C13	Total MS/MS intensity
FFV_EP_22.d, MS/MS of 345.1363085 1+ at 1.75028333333333 mins	1.75	345.136	345.1375	-4.3	2_0_0_0_0	Hex_2 RED_1 H2O_1	1	0.77	1	0	13188.17826
FFV_EP_22.d, MS/MS of 315.1276143 1+ at 1.76548333333333 mins	1.765	315.128	315.127	3.2	1_1_0_0_0	Hex_1 Pent_1 RED_1 H2O_1	1	0.85	1	0	1358.20148
FFV_EP_22.d, MS/MS of 507.1895569 1+ at 2.84265 mins	2.843	507.19	507.1903	-0.6	3_0_0_0_0	Hex_3 RED_1 H2O_1	1	0.76	1	0	794.63192
FFV_EP_22.d, MS/MS of 507.1895569 1+ at 4.15603333333333 mins	4.156	507.19	507.1903	-0.6	3_0_0_0_0	Hex_3 RED_1 H2O_1	1	0.82	1	0	2813.2559
FFV_EP_22.d, MS/MS of 507.1895569 1+ at 4.34954973958333 mins	4.35	507.19	507.1903	-0.6	3_0_0_0_0	Hex_3 RED_1 H2O_1	1	0.84	1	0	2269.89584
FFV_EP_22.d, MS/MS of 477.1779683 1+ at 4.48158333333333 mins	4.482	477.178	477.1798	-3.8	2_1_0_0_0	Hex_2 Pent_1 RED_1 H2O_1	1	0.85	1	0	16139.21488
FFV_EP_22.d, MS/MS of 345.1363085 1+ at 4.6077 mins	4.608	345.136	345.1375	-4.3	2_0_0_0_0	Hex_2 RED_1 H2O_1	1	0.8	1	0	304.25362
FFV_EP_22.d, MS/MS of 477.1779683 1+ at 4.61286666666667 mins	4.613	477.178	477.1798	-3.8	2_1_0_0_0	Hex_2 Pent_1 RED_1 H2O_1	1	0.82	1	0	7908.16207
FFV_EP_22.d, MS/MS of 507.1895569 1+ at 4.69048333333333 mins	4.69	507.19	507.1903	-0.6	3_0_0_0_0	Hex_3 RED_1 H2O_1	1	0.86	1	0	9821.95729
FFV_EP_22.d, MS/MS of 507.1895569 1+ at 4.79488333333333 mins	4.795	507.19	507.1903	-0.6	3_0_0_0_0	Hex_3 RED_1 H2O_1	1	0.8	1	0	4742.05885
FFV_EP_22.d, MS/MS of 669.2412821 1+ at 8.33273333333333 mins	8.333	669.241	669.2431	-3.1	4_0_0_0_0	Hex_4 RED_1 H2O_1	1	0.91	1	0	996.49685
FFV_EP_22.d, MS/MS of 669.2412821 1+ at 9.19586666666667 mins	9.196	669.241	669.2431	-3.1	4_0_0_0_0	Hex_4 RED_1 H2O_1	1	0.82	1	0	1085.19717
FFV_EP_22.d, MS/MS of 669.2412821 1+ at 9.38093333333333 mins	9.381	669.241	669.2431	-3.1	4_0_0_0_0	Hex_4 RED_1 H2O_1	1	0.95	1	0	318.94732
FFV_EP_22.d, MS/MS of 683.2210693 1+ at 10.9668166666667 mins	10.967	683.221	683.2224	-2	3_0_1_0_0	Hex_3 HexA_1 RED_1 H2O_1	1	0.79	1	0	1509.48571
FFV_EP_22.d, MS/MS of 683.2210693 1+ at 11.0729666666667 mins	11.073	683.221	683.2224	-2	3_0_1_0_0	Hex_3 HexA_1 RED_1 H2O_1	1	0.82	1	0	1017.6058
FFV_EP_22.d, MS/MS of 669.2412821 1+ at 11.4957333333333 mins	11.496	669.241	669.2431	-3.1	4_0_0_0_0	Hex_4 RED_1 H2O_1	1	0.84	1	0	3007.29538
FFV_EP_22.d, MS/MS of 669.2412821 1+ at 12.3089 mins	12.309	669.241	669.2431	-3.1	4_0_0_0_0	Hex_4 RED_1 H2O_1	1	0.85	1	0	8415.76539
FFV_EP_22.d, MS/MS of 639.2306671 1+ at 12.31455 mins	12.315	639.231	639.2326	-2.5	3_1_0_0_0	Hex_3 Pent_1 RED_1 H2O_1	1	0.86	1	0	6751.11148
FFV_EP_22.d, MS/MS of 507.1895569 1+ at 12.40875 mins	12.409	507.19	507.1903	-0.6	3_0_0_0_0	Hex_3 RED_1 H2O_1	1	0.86	0.5	0	344.49098
FFV_EP_22.d, MS/MS of 683.2210693 1+ at 13.71605 mins	13.716	683.221	683.2224	-2	3_0_1_0_0	Hex_3 HexA_1 RED_1 H2O_1	1	0.78	1	0	3468.78336
FFV_EP_22.d, MS/MS of 653.2086792 1+ at 13.7321166666667 mins	13.732	653.209	653.2119	-4.4	2_1_1_0_0	Hex_2 Pent_1 HexA_1 RED_1 H2O_1	1	0.78	1	0	4431.91395

FFV_EP_22.d, MS/MS of 831.2917786 1+ at 14.2426666666667 mins	14.243	831.292	831.2959	-4.7	5_0_0_0_0	Hex_5 RED_1 H2O_1	1	0.85	1	0	10996.759
FFV_EP_22.d, MS/MS of 801.2846273 1+ at 14.31205 mins	14.312	801.285	801.2854	-0.5	4_1_0_0_0	Hex_4 Pent_1 RED_1 H2O_1	1	0.85	1	0	13952.55848
FFV_EP_22.d, MS/MS of 801.2846273 1+ at 14.39795 mins	14.398	801.285	801.2854	-0.5	4_1_0_0_0	Hex_4 Pent_1 RED_1 H2O_1	1	0.77	1	0	2529.5005
FFV_EP_22.d, MS/MS of 993.3475138 1+ at 14.86645 mins	14.866	993.348	993.3487	-0.7	6_0_0_0_0	Hex_6 RED_1 H2O_1	1	0.78	0.8	0	329.93441
FFV_EP_22.d, MS/MS of 845.2740479 1+ at 14.9462833333333 mins	14.946	845.274	845.2752	-1.4	4_0_1_0_0	Hex_4 HexA_1 RED_1 H2O_1	1	0.78	1	0	2575.40654
FFV_EP_22.d, MS/MS of 993.3475138 1+ at 15.1956666666667 mins	15.196	993.348	993.3487	-0.7	6_0_0_0_0	Hex_6 RED_1 H2O_1	1	0.79	1	0	3403.22219
FFV_EP_22.d, MS/MS of 963.3362732 1+ at 15.3622333333333 mins	15.362	963.336	963.3382	-2.3	5_1_0_0_0	Hex_5 Pent_1 RED_1 H2O_1	1	0.81	1	0	9979.86003
FFV_EP_22.d, MS/MS of 507.1895569 1+ at 15.5079 mins	15.508	507.19	507.1903	-0.6	3_0_0_0_0	Hex_3 RED_1 H2O_1	1	0.81	1	0	12628.54311
FFV_EP_22.d, MS/MS of 845.2740479 1+ at 15.5342166666667 mins	15.534	845.274	845.2752	-1.4	4_0_1_0_0	Hex_4 HexA_1 RED_1 H2O_1	1	0.83	1	0	9103.93415
FFV_EP_22.d, MS/MS of 507.1895569 1+ at 15.6439833333333 mins	15.644	507.19	507.1903	-0.6	3_0_0_0_0	Hex_3 RED_1 H2O_1	1	0.76	1	0	790.19129
FFV_EP_22.d, MS/MS of 1155.3984985 1+ at 16.0046 mins	16.005	1155.398	1155.4015	-3	7_0_0_0_0	Hex_7 RED_1 H2O_1	1	0.77	1	0	12772.87013
FFV_EP_22.d, MS/MS of 959.3047485 1+ at 16.3166166666667 mins	16.317	959.305	959.307	-2.1	1_1_2_2_0	Hex_1 Pent_1 HexA_2 dHex_2 RED_1 H2O_1	1	0.82	1	0	986.57591
FFV_EP_22.d, MS/MS of 959.3047485 1+ at 16.3166166666667 mins	16.317	959.305	959.3071	-2.2	1_2_2_1_1	Hex_1 Pent_2 HexA_2 dHex_1 CH3_1 RED_1 H2O_1	1	0.82	0.83	0	986.57591
FFV_EP_22.d, MS/MS of 815.2635345 1+ at 16.4490833333333 mins	16.449	815.264	815.2647	-0.9	3_1_1_0_0	Hex_3 Pent_1 HexA_1 RED_1 H2O_1	1	0.79	1	0	2139.97858
FFV_EP_22.d, MS/MS of 1007.3240153 1+ at 16.6341666666667 mins	16.634	1007.324	1007.328	-4	5_0_1_0_0	Hex_5 HexA_1 RED_1 H2O_1	1	0.81	1	0	4980.36813
FFV_EP_22.d, MS/MS of 1121.3542480 1+ at 16.7861666666667 mins	16.786	1121.354	1121.3598	-5.2	2_1_2_2_0	Hex_2 Pent_1 HexA_2 dHex_2 RED_1 H2O_1	1	0.78	1	0	2026.1931
FFV_EP_22.d, MS/MS of 1121.3542480 1+ at 16.7861666666667 mins	16.786	1121.354	1121.3599	-5.3	2_2_2_1_1	Hex_2 Pent_2 HexA_2 dHex_1 CH3_1 RED_1 H2O_1	1	0.78	1	0	2026.1931
FFV_EP_22.d, MS/MS of 983.3353271 2+ at 21.4182333333333 mins	21.418	1965.6627	1965.6655	-1.4	12_0_0_0_0	Hex_12 RED_1 H2O_1	2	0.81	0.54	0	203.79079
FFV_EP_22.d, MS/MS of 831.2917786 1+ at 27.1899666666667 mins	27.19	831.292	831.2959	-4.7	5_0_0_0_0	Hex_5 RED_1 H2O_1	1	0.8	1	0	5312.20262
FFV_EP_22.d, MS/MS of 831.2917786 1+ at 27.3742 mins	27.374	831.292	831.2959	-4.7	5_0_0_0_0	Hex_5 RED_1 H2O_1	1	0.81	1	0	761.0061

For the label-free relative quantitation, oligosaccharides were first assigned to their parent polysaccharide using the fingerprint library. Peak areas from the MS1 chromatograms of these oligosaccharides were summed for each polysaccharide. Relative abundances were then derived from the normalized peak area sum for each sample. Label-free relative quantitation is usually employed in other LC-MS-based -omics methodologies because of its ease and simplicity.[25,26] MS1-based relative quantitation can be used to compare across samples, but within-sample comparisons are generally not accurate due to differences in the ionization of different compounds.

Absolute quantitation using external calibration curves were done in Microsoft Excel. For each polysaccharide, peak areas of the top 3 most abundant oligosaccharides were averaged and used for the calibration curve. The range of the calibration curve varied between different polysaccharides. The highest calibrator ranged from at least 1.00 to 8.50 mg/mL and then these were serially diluted as follows: 2×, 4×, 8×, 40×, 80×. At least five points were used in the linear regression fit (equal weighing) and the intercepts were forced to zero.

Further statistics and visualization were done with R programming language. For the absolute quantitation of apples and onions, multiple one-way analysis of variance (ANOVA) was done for each polysaccharide. *P*-values were adjusted for multiple testing using Benjamini-Hochberg (also called false-discovery rate adjustment or FDR-adjusted). Pairwise post hoc mean comparisons were done using Tukey's test.

Results and Discussion

By using a non-enzymatic bottom-up approach for polysaccharide analysis, we have developed a high-throughput, multiplexed, and quantitative method to analyze polysaccharides in food samples. Multiplexing was enabled by the FITDOG reaction in which multiple polysaccharides with diverse chemical structures were depolymerized into distinct oligosaccharides products. Polysaccharides standards were reacted using FITDOG and the oligosaccharides were used to construct a fingerprint library.[18,19] By using external calibration curves, we have further extended the application to absolute quantitation in the more complex food samples, chosen to contain different types and amounts of polysaccharides. Quantitation of polysaccharides using the proposed methodology (**Figure 3.1**) was validated for recovery using the commercially available polysaccharide standards.

Liquid chromatography using porous graphitic carbon (PGC)

The retention mechanism of PGC is influenced mainly by the planarity and hydrophilicity of the analytes.[22] PGC had been previously shown to be very effective in resolving isomeric carbohydrate structures and so it is the most appropriate separation column for this workflow.[23,24] The capability of PGC to resolve isomeric oligosaccharides was shown with Hex_5 oligosaccharides (**Table 3.3**). The different Hex_5 oligosaccharides have varying monosaccharide and linkage compositions and they were all resolved in the chromatographic dimension. For example, cellulose oligosaccharides with β(1→4)-Glc linkage have more planar structure than amylose[1], and therefore were retained more in the PGC column, as demonstrated by higher retention time values. Among the Hex_5 oligosaccharides with β(1→4) linkage listed in **Table**

3.3, mannopentaose and galactopentaose oligosaccharides were less retained than cellopentaose, most probably due to the axial hydroxyl group orientations at C2 and C4 positions in mannose and galactose, respectively.

Table 3.3. Example of isomeric separation of Hex5 isomers using porous graphitic carbon (PGC) as analytical column.

Composition	RT (min)	Linkage [1]	Polysaccharide	Chromatographic Resolution
Hex_5	10.25	β(1→4)Man	mannan	--
Hex_5	11.75	β(1→4)Gal	galactan	2
Hex_5	12.9	α(1→4)Glc	amylose	1.53
Hex_5	26.7	β(1→4)Glc	cellulose	18.4
Hex_5	27.5	β(1→3)Glc	curdlan	1.07

[1] Linkages were inferred from known structures of the parent polysaccharides.

Generation of fingerprint profile for the polysaccharides

Example oligosaccharide chromatograms from FITDOG-reacted polysaccharide standards are shown in **Figure 3.3**. Starch is comprised of amylose and amylopectin, where amylose is a linear homopolymer of glucose connected with α(1→4) linkage, while amylopectin is similar to amylose with branching points with α(1→6) linkage.[1] The FITDOG reaction with starch yielded oligosaccharides of varying degrees of polymerization (DP), ranging from 3 up to 21. Both amylose and amylopectin standards gave similar oligosaccharide profiles after reaction with FITDOG. Cellulose is another linear homopolymer of glucose connected with β(1→4) linkage. The difference in anomeric configuration between starch and cellulose oligosaccharides resulted in distinct oligosaccharide profiles. Galactan polysaccharide is comprised of β(1→4)-

linked galactose residues and is usually attached as a side branch in pectin polysaccharides.[21] Galactan oligosaccharides resulting from the FITDOG reaction ranged from DP 3 up to DP 12. Arabinan is another domain present in pectin polysaccharides, where the backbone is comprised of α(1→5)-arabinofuranose residues.[22] Xylan is a plant polysaccharide with a linear backbone of β(1→4)-xylose and occasionally with branches of glucuronic acid residues.[23,24] The glucuronic acid residues are further typically *O*-methylated at the C4 position. Xylan oligosaccharides, including the methylated glucuronic acid residues, were detected using the FITDOG workflow. Oligosaccharide chromatogram profiles of other polysaccharides (mannan, chitin, β-glucan, xyloglucan) were shown in **Figure 3.4**. The complete oligosaccharide fingerprint library is summarized in **Table 3.1**.

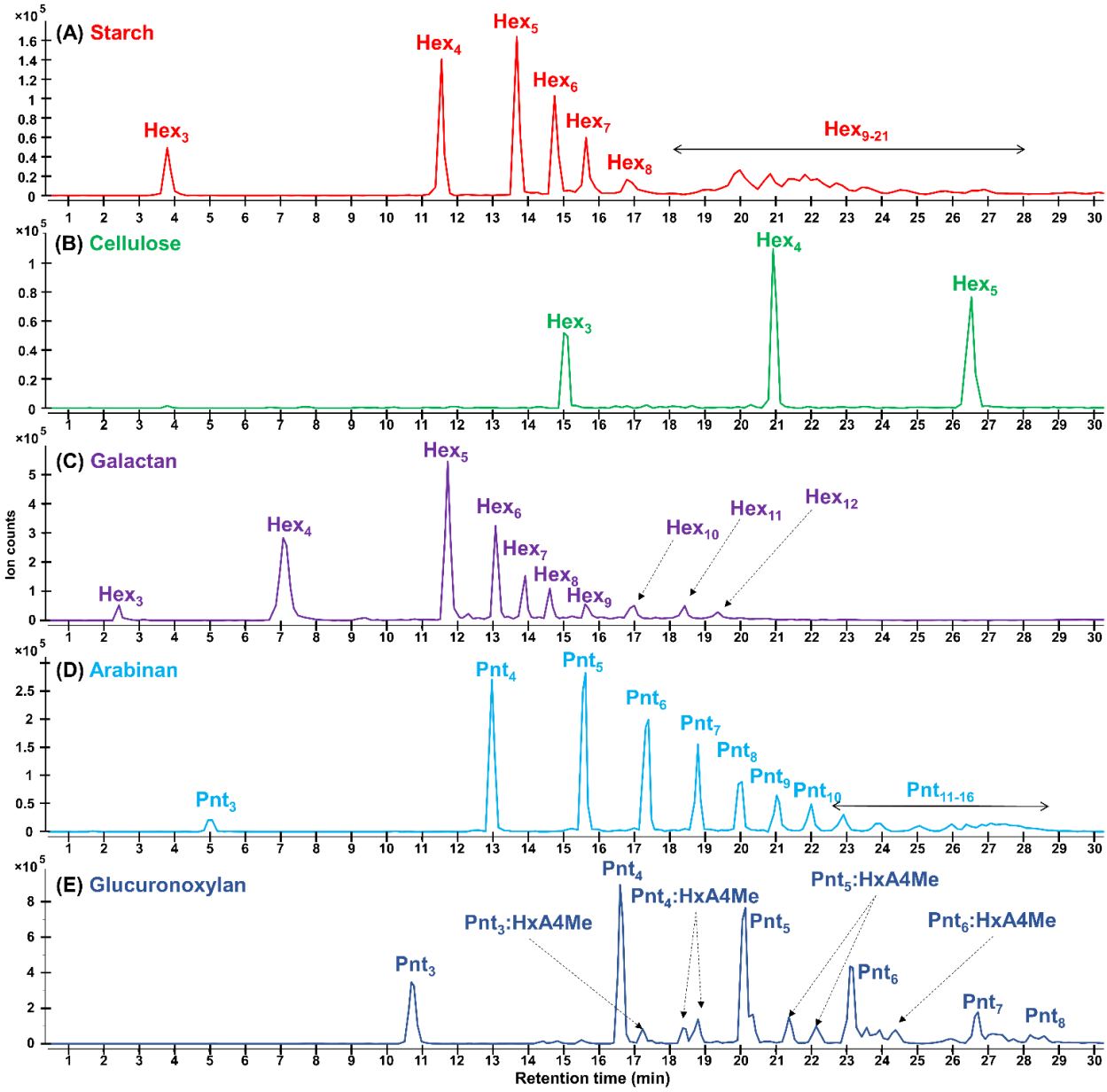


Figure 3.3. Example chromatograms showing oligosaccharide products from FITDOG reactions of each polysaccharide. Hex = hexose, Pnt = pentose, HxA = hexuronic acid, 4Me = 4-O-methyl.

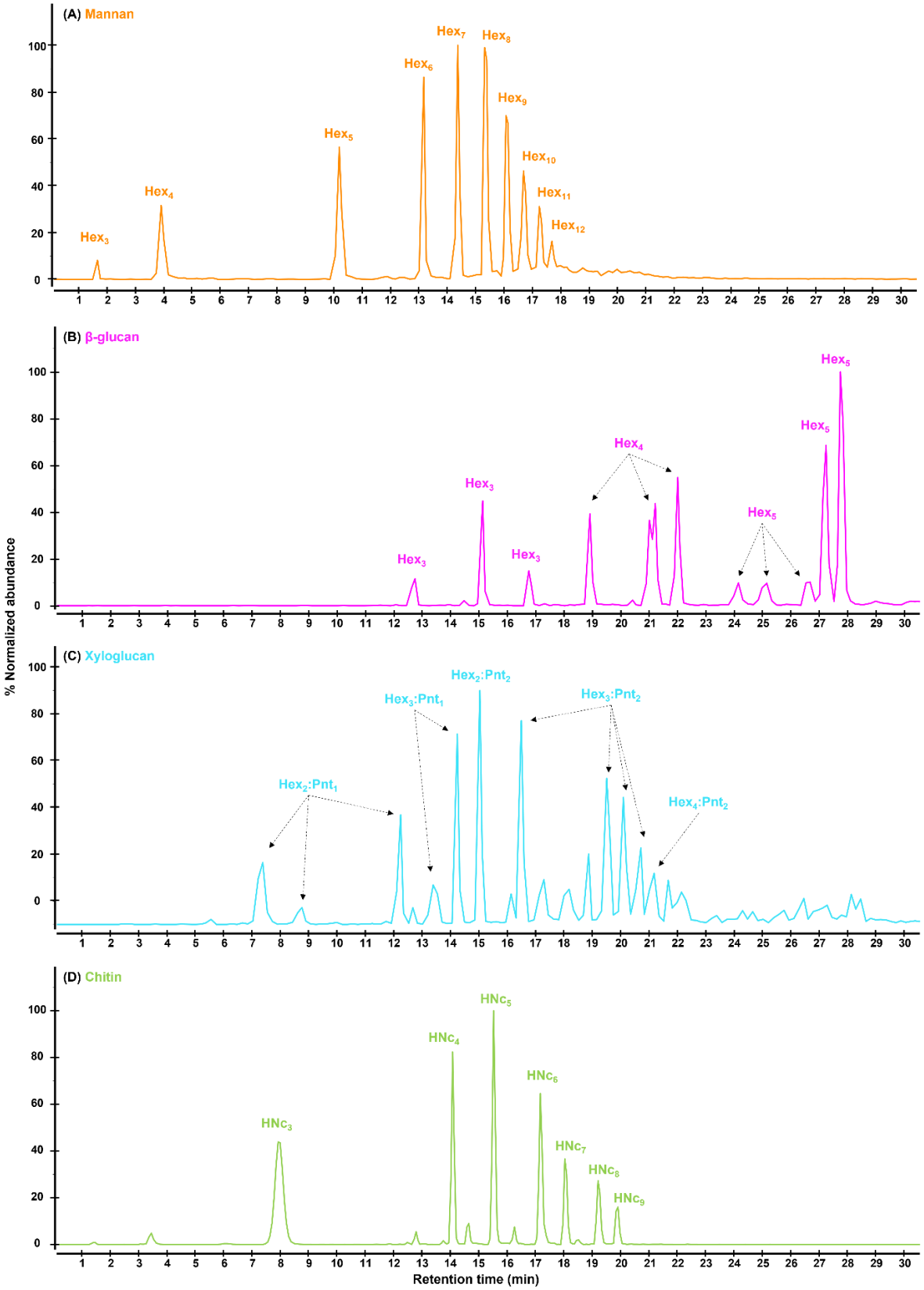


Figure 3.4. Chromatogram profiles of oligosaccharide products from polysaccharide standards reacted with FITDOG. (**A**) mannan, (**B**) β-glucan, (**C**) xyloglucan, and (**D**) chitin.

Validation of quantitation using oligosaccharide and polysaccharide standards

Quantitative results were validated by using commercially available standards. First, commercial oligosaccharide standards were pooled and serially diluted at different concentrations and injected in the HPLC-QTOF to determine the instrument response with respect to concentrations (**Figure 3.5**). This demonstrated that the HPLC-QTOF method generated proportional changes in the peak area in response to analyte concentration and can be amenable to quantitation. Different compounds gave distinct relative responses as measured by the slopes of the fitted linear regression. This observation highlighted the need for generating a separate calibration curve for each analyte of interest.

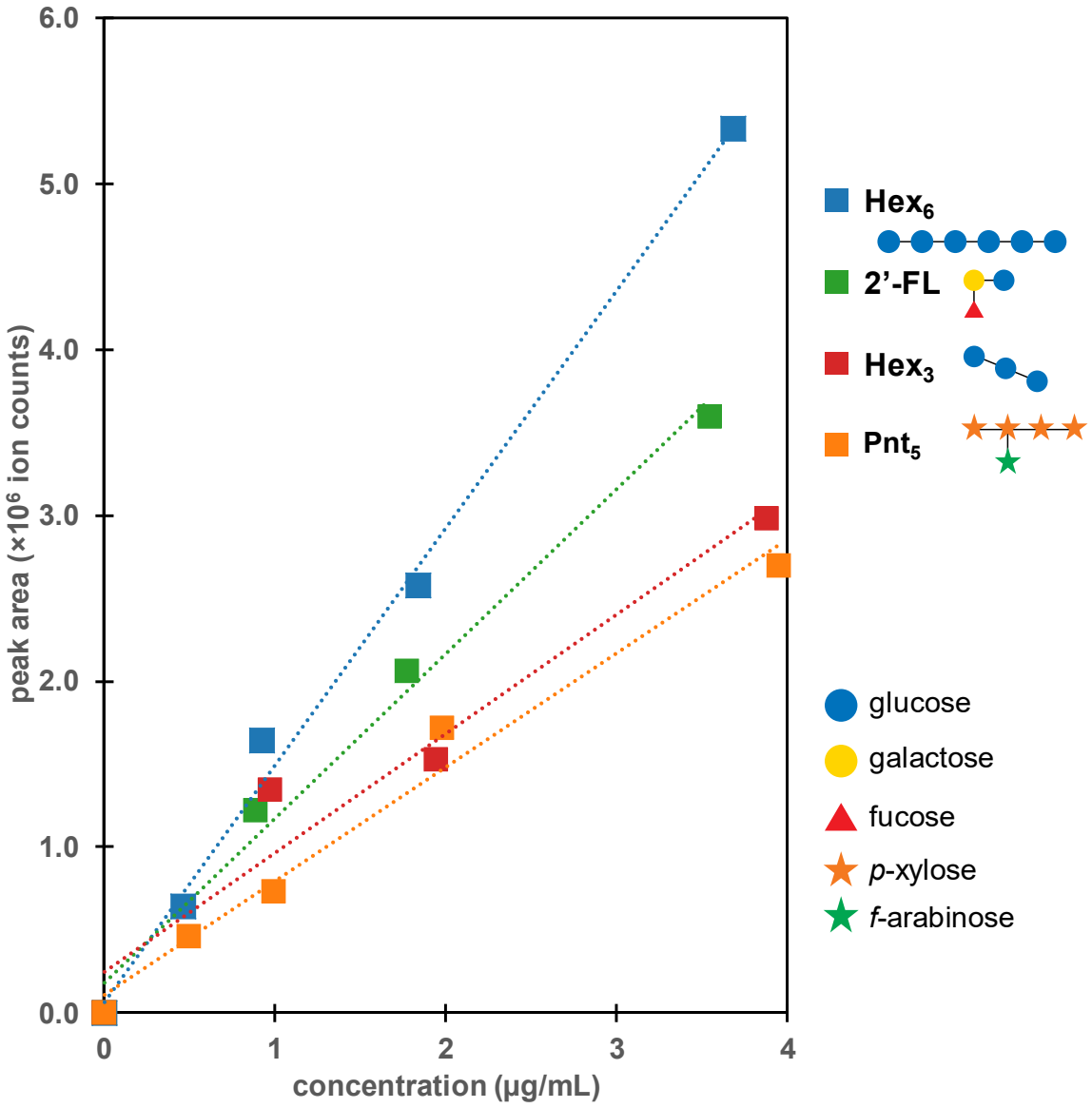


Figure 3.5. Peak areas of oligosaccharides (Hex_6 = maltohexaose, 2'-FL = 2'-fucosyllactose, Hex_3 = isomaltotriose, Pnt_5 = 3^3-α-L-arabinofuranosyl-xylotetraose) plotted against injection concentration. Monosaccharide legend: blue circle = glucose, yellow circle = galactose, red triangle = fucose, orange star = xylose, green pentagram = arabinose.

Quantitation of polysaccharides was evaluated using calibration curves prepared by subjecting polysaccharide standards (starch, cellulose, β-glucan, xyloglucan, mannan, galactan, arabinan, xylan, chitin) to the FITDOG analysis. Several pooled mixtures of polysaccharide

standards were prepared and serially diluted to generate the calibration curve standards. To get a more representative quantitation metric, chromatographic peak areas of the top three most abundant unique oligosaccharides from each polysaccharide were averaged and was used for the calibration curves. For example, the linear arabinan standard yielded 13 oligosaccharides that could be used for quantitation. From these arabinan oligosaccharides, Arb_3, Arb_4, and Arb_5 were the most abundant and their peak areas were averaged and used for the calibration curve. Representative calibration curves for the linear arabinan standard and for starch are shown in **Figures 3.6A** and **3.6B**. This calibration process was done for all the other polysaccharides (**Table 3.4, Figure 3.7**). Overall, most calibration curves were linear ($r^2 > 0.99$), except for chitin ($r^2 = 0.98$). Among the polysaccharides, chitin had the highest slope while cellulose had the lowest slope. The method detection limit (MDL) was estimated based on the lowest concentration of standard reacted which gave an averaged peak area signal-to-noise ratio (S/R) value > 3. Chitin and arabinan had the lowest MDL (~55 μg/mL or ~0.22% wt/wt dry basis). The linear ranges spanned approximately two orders of magnitude for all polysaccharides.

Table 3.4. Calibration curve parameters for the absolute quantitation of polysaccharides using the quantitative FITDOG method.

Polysaccharide	**r^2**	**Slope**	**MDL (μg/mL)**	**MDL (%wt/wt, dry)**	**S/R @ MDL**
β-glucan	0.999	55325	96	0.38	2.7
chitin	0.979	117476	55	0.22	6.1
mannan	0.995	35637	89	0.36	5.1
xylan	0.995	51848	103	0.41	10.0
arabinan	0.998	69723	57	0.23	5.3
galactan	0.996	24987	532	2.13	8.0
xyloglucan	0.997	26912	542	2.17	7.5
cellulose	0.997	9783	350	1.40	3.7
starch	0.999	34165	538	2.15	16.6

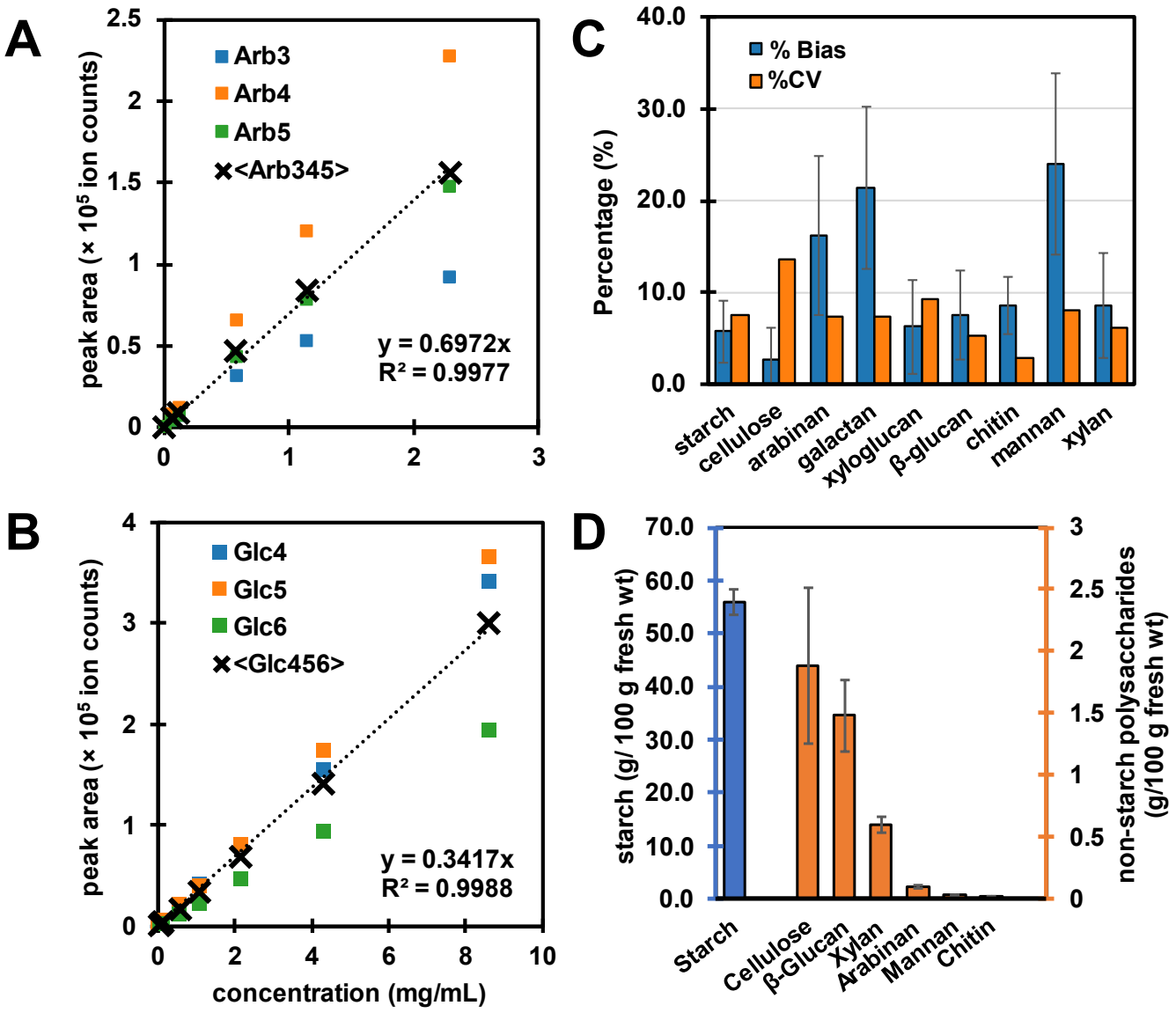


Figure 3.6. Quantitative results from FITDOG analysis polysaccharide standards. External calibration curves for (**A**) linear arabinan standard and (**B**) starch. Arb = arabinose, Glc = glucose. (**C**) Accuracy (% error) and reproducibility (% CV) based on mixtures of polysaccharide standards. %CV was computed based on 3 method replicates. (**D**) FITDOG results for polysaccharides in a standard reference material (Fortified Breakfast Cereal, NIST SRM® 3233). Right vertical axis corresponds to non-starch polysaccharides (chitin, mannan, arabinan, xylan, β-glucan, cellulose), while left vertical axis corresponds to starch values.

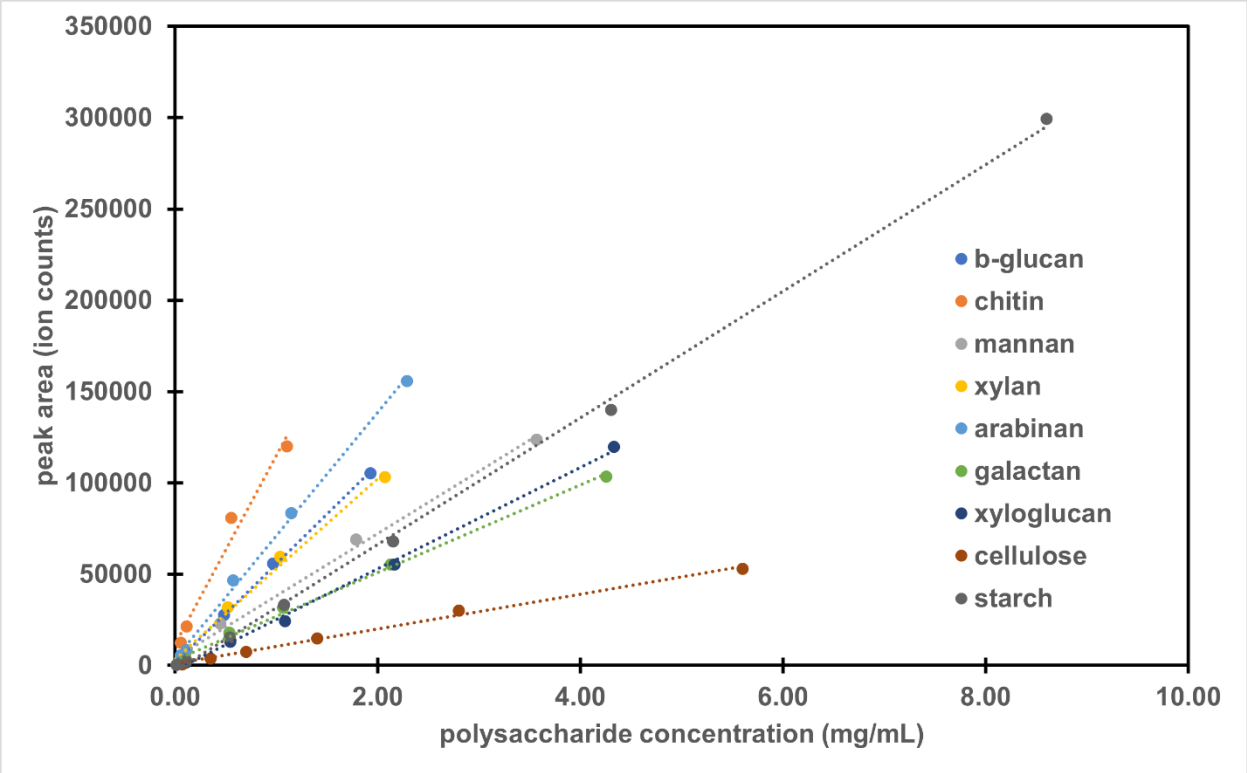


Figure 3.7. Examples of calibration curves of polysaccharide using the quantitative workflow presented in the paper.

To verify the quantitative approach, several pooled mixtures of standard polysaccharides were prepared, analyzed, and quantified using the proposed calibration method. The accuracy of the method was quantified by percent (%) difference between the measured and expected concentration (based on nominal concentration of the test mixtures), while the reproducibility was demonstrated by % coefficient of variation (CV) based on 3 technical replicates taken through the entire method (**Figure 3.6C**). The accuracy ranged from 5% to 25% bias, while the reproducibility ranged from 2% to 15% CV. In terms of accuracy, arabinan, galactan, and mannan values had the most deviation (>15%) from the expected concentration, while starch, β-glucan, xyloglucan, and xylan had the least deviations (<10%). Six out of nine polysaccharides

quantified had <10% error. Furthermore, the workflow was highly reproducible with %CV of less than 10% for all polysaccharides except cellulose (14% CV).

To test method performance on a food matrix, NIST SRM® 3233 Fortified Breakfast Cereal (National Institute of Standards and Technology, Gaithersburg, MD, USA) was analyzed, in triplicate (**Figure 3.6D**). The Certificate of Analysis (COA) for this material includes reference values for total carbohydrates by difference (79.23 ± 1.04 g/100 g), total free sugars (16.07 ± 1.53 g/100 g), and low molecular weight soluble dietary fiber (LMW SDF, 3.07 ± 0.62 g/100 g). Total carbohydrates by difference include all forms of carbohydrates, including free sugars, oligosaccharides, and all polysaccharides. Subtracting total free sugars and the LMW SDF fraction from the total carbohydrates will provide an estimate of the polysaccharide fraction of the cereal standard.[25] Total assayed polysaccharides from FITDOG (sum of starch, cellulose, mannan, β-glucan, chitin, arabinan, xylan) was 60.02 ± 2.63 g/100 g and this was within the expected range for total polysaccharide estimated from COA values (**Table 3.5**).

Table 3.5. Summary of results for Fortified Breakfast Cereal (NIST SRM® 3233) based on Certificate of Analysis (COA) and the FITDOG method.

COA				**FITDOG**			
attribute	**g/ 100 g fresh wt**			**polysaccharide**	**g/ 100 g fresh wt**		
total carbohydrates	79.23	±	1.04	Starch	55.92	±	2.54
total sugars	16.07	±	1.53	Cellulose	1.88	±	0.63
LMW SDF	3.07	±	0.62	β-Glucan	1.48	±	0.29
				Xylan	0.59	±	0.07
				Arabinan	0.10	±	0.02
				Mannan	0.03	±	0.01
				Chitin	0.01	±	0.01
Estimated polysaccharide	**60.08**	**±**	**1.95**	**Total polysaccharide**	**60.02**	**±**	**2.63**

To assess the reproducibility of the various steps of the workflow, five commercial polysaccharide standards were pooled, reacted, and injected to the instrument. Each step of the workflow was done with 6-7 replicates and at least 29 oligosaccharides were monitored from the 5 polysaccharides (**Figure 3.8**). In this experiment, samples from the previous step of the method were aliquoted and pooled to serve as the replicates for the next step. The largest variations were observed with replicates taken through the entire assay starting from the FITDOG reaction step. Less variations were observed from replications in the subsequent steps of the workflow, namely in the $NaBH_4$ reduction and SPE clean-up. Overall, the validation experiments using standards demonstrated the accuracy and reproducibility of the FITDOG workflow for multiplexed, high-throughput, absolute quantitation of polysaccharides.

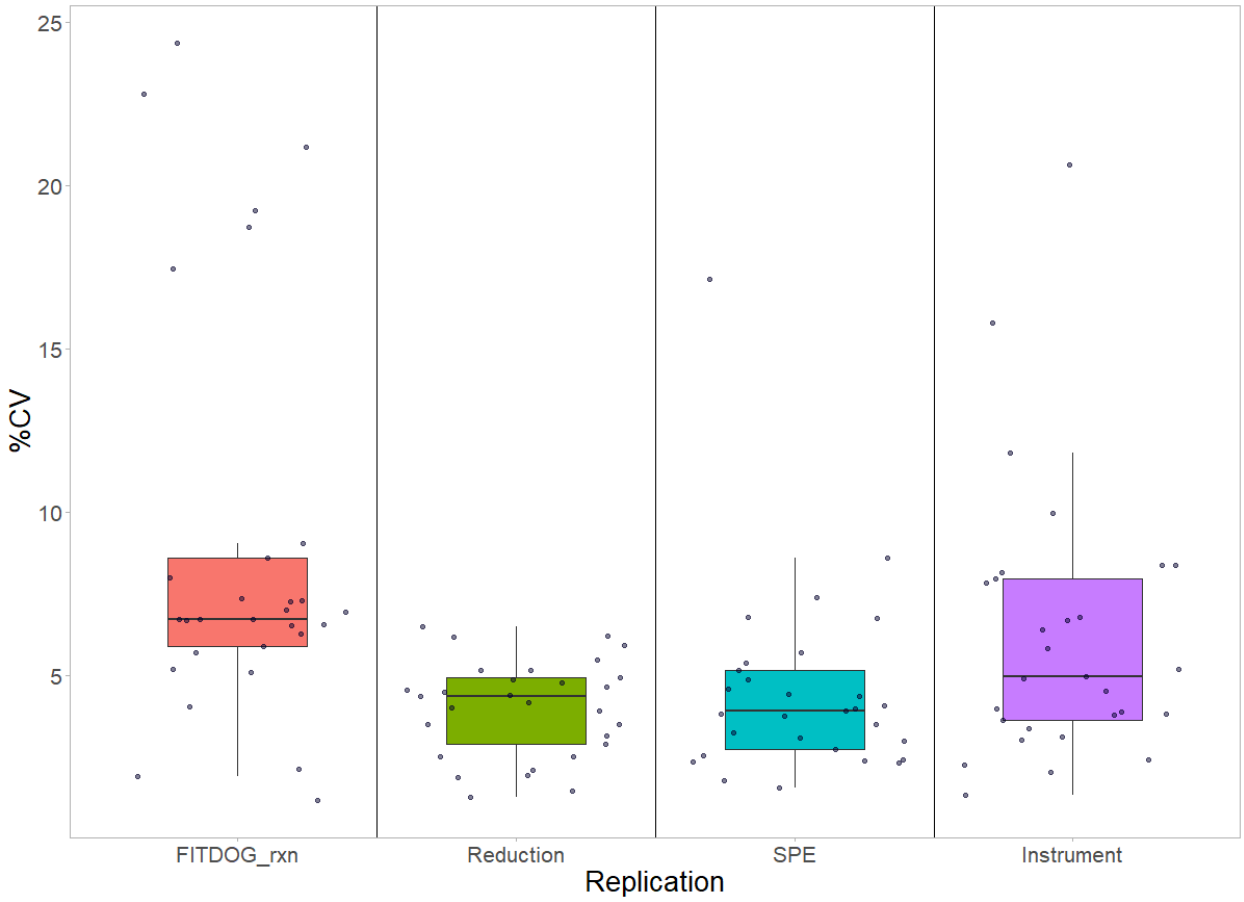


Figure 3.8. Reproducibility for each step of the quantitative analysis of polysaccharides. Each point corresponds to an oligosaccharide and each step had 6-7 replicates.

Quantitation of polysaccharides in food samples

Representative chromatograms for select food samples are shown in **Figure 3.9**. In both artichoke samples (**Figures 3.9A, 3.9B**), the oligosaccharide fingerprints showed cellulose and xylan, but starch was only in the inner leaves sample. The avocado seed (**Figure 3.9C**) showed high amount of starch, while the avocado skin (**Figure 3.9D**) had glucuronoxylan and cellulose. Avocado seed has been previously shown to contain high amounts of starch.[26]

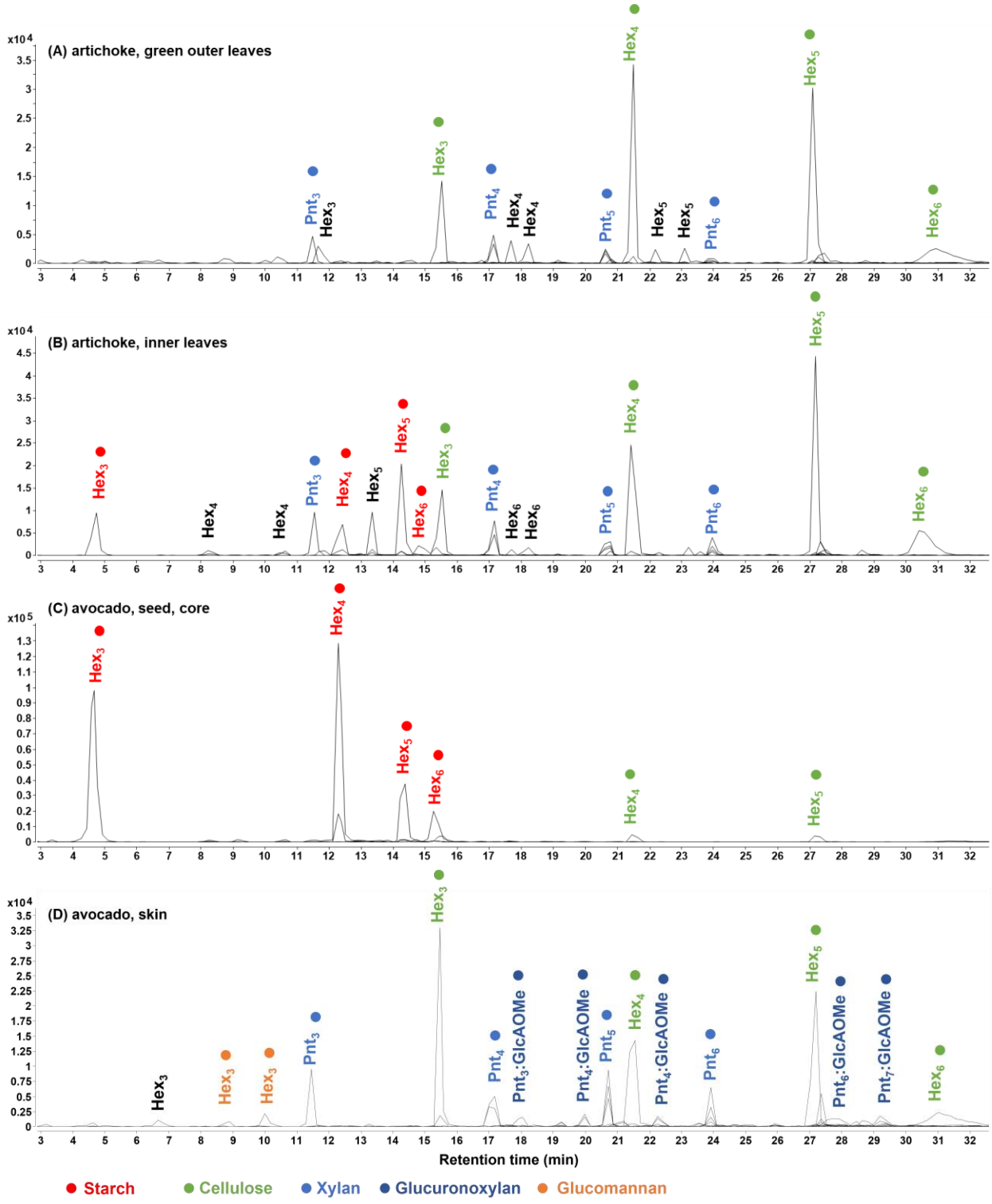


Figure 3.9. Examples of annotated chromatograms (oligosaccharide profiles) of various food samples. **(A)** green outer leaves from artichoke, **(B)** inner leaves from artichoke, **(C)** avocado seed, and (**D**) avocado skin. Y-axes are in ion counts.

Relative quantitation results for single measurements of 13 different foods are shown in **Figure 3.10**. To get a diverse set of polysaccharides and to demonstrate the generality of the method, these foods were also partitioned to several anatomical parts, including some non-edible parts, for analysis. The relative abundance was based on peak areas of extracted ion chromatograms of the oligosaccharides resulting from the FITDOG reaction. Broccoli stems and green onion had the highest relative amount of cellulose. Avocado seed had the highest amount of starch out of all the samples analyzed. Okra and some artichoke parts showed appreciable amounts of starch. Xylans were also detected in lower amounts in several artichoke samples, avocado skin, and sage stem. Unassigned oligosaccharides referred to peaks identified as oligosaccharides based on tandem mass spectra but were not matched to any polysaccharide based on retention times from the oligosaccharide fingerprinting library. These unassigned oligosaccharides accounted for 20-30% relative abundance based on peak area across all samples and were mostly Hex_n:Pnt_1 and Hex_n:Pnt_1:HxA_1 oligosaccharides.

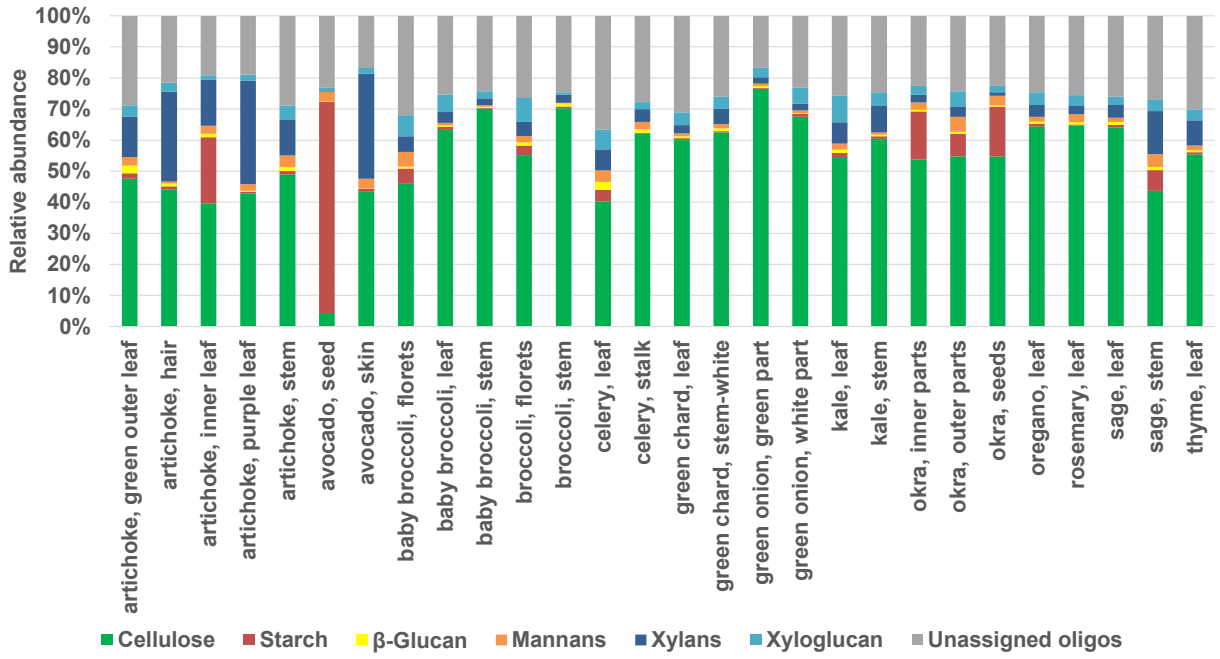


Figure 3.10. Relative quantitation of polysaccharides in food samples based on extracted ion chromatogram peak area abundances.

Finally, using the external calibration curve, the absolute quantitation workflow was then applied to several sample sets consisting of apples and onions (**Figure 3.11**). The apple set consisted of five varieties, where each had 7-8 samples obtained from different sources. Three varieties of onions were analyzed, each with 6 samples obtained from different sources.

Among the apples, cellulose was the most abundant polysaccharide which ranged from 1 to 2% (wt/wt by fresh weight). Galactans, arabinans, and starch had values < 0.2%. Based on analysis of variance (ANOVA) done for each analyte, the five most abundant polysaccharides (cellulose, arabinan, galactan, mannan, and starch) were statistically significant between the five varieties analyzed (FDR-adjusted $p < 0.05$). Additionally, *post hoc* comparison tests using Tukey's test revealed some differences between these five varieties. For example, Granny Smith apple had significantly higher amounts of galactan, arabinan, and mannan compared to the other

4 varieties. Red Delicious apple had the highest amount of starch, while Fuji and Honeycrisp had the lowest starch content. Although cellulose had a significant *p*-value from the ANOVA, the *post hoc* comparison tests did not show any significant pairwise comparisons between the five varieties. It has been previously shown that apples have significant amounts of cellulose.[27] Arabinans and galactans were observed at appreciable amounts in the apples analyzed, which was expected based on previously published data on the carbohydrate characterization of apples. Pectic polysaccharides in apples are known to contain both arabinan and galactan branches.[28]

Among the onions, no significant differences were found across all polysaccharides between the three varieties analyzed. White onion had consistently lower amounts of polysaccharides. Cellulose was again the most abundant polysaccharide, followed by galactan and starch. Previously published data has shown that onion cell walls are comprised mostly of pectins, hemicelluloses, and cellulose.[29,30]

The apple and onion samples demonstrated that results from our workflow corroborated with existing data on the expected polysaccharides found in these samples. However, the quality of quantitative data obtained from our workflow is unprecedented in terms of scale, coverage, and throughput. Previously published papers on food carbohydrate analysis involved complex fractionation schemes, and quantitation from these studies is often limited to monosaccharide and glycosidic linkage analyses.

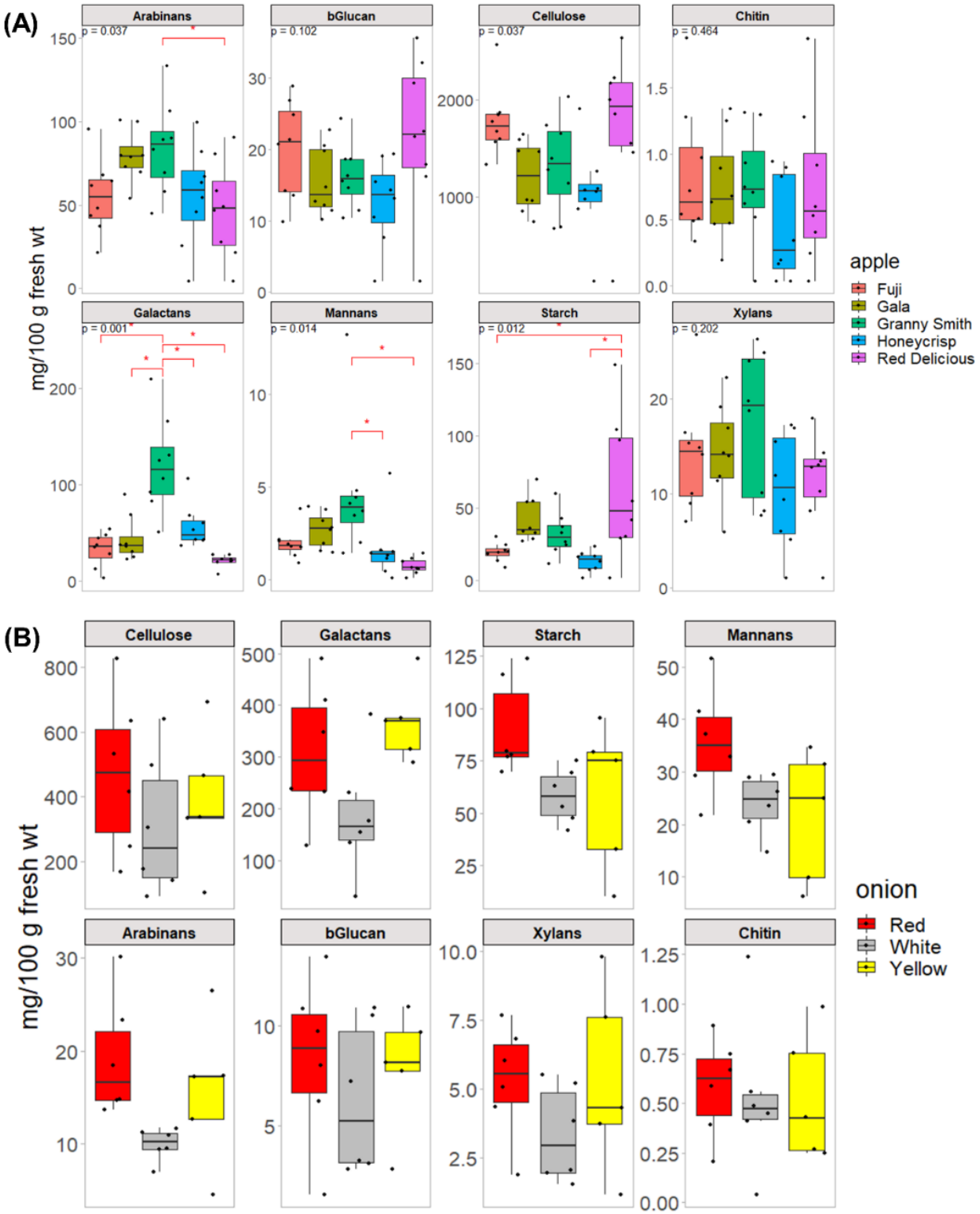


Figure 3.11. Boxplot graphs showing content of different polysaccharide types among varieties of (**A**) apples (7-8 independent retail samples of each of Fuji, Gala, Granny Smith, Honeycrisp, Red Delicious varieties), and (**B**) onions (6 independent retail samples of each of white, red, and yellow varieties). *P*-values shown in (**A**) were false-discovery-rate adjusted *p*-values from ANOVA, while * denotes *post hoc* Tukey's test $p < 0.05$.

Limitations and future work

The reported approach has not been optimized yet to detect and quantitate some other food polysaccharides, such as fructans (e.g., inulin, levan) and other pectic polysaccharides, such as polygalacturonans, and rhamnogalacturonans. Galacturonans are anionic polysaccharides containing galacturonic acid residues.[1] These anionic oligosaccharides can potentially be analyzed better in negative-mode ionization. Additionally, as discussed in the text before, the recoveries for some polysaccharides could still be improved. Standard addition method can be used however, throughput will slightly decrease due to the number of samples necessary for standard addition. Using internal standards can be further explored, although stable-isotope-labelled polysaccharides are generally uncommon and can be prohibitively expensive. Nevertheless, we envision that this approach will lead to new techniques to be developed to analyze polysaccharides, especially due to its compatibility with being conducted in a multiplexed, high-throughput semi-automated workflow.

Conclusion

A high-throughput method enabling accurate and reproducible qualitative and quantitative characterization of polysaccharides in food samples was successfully demonstrated using a bottom-up glycomics approach. The method is suitable for quantitation of common food polysaccharides (e.g. starch, cellulose, mannans, arabinans, xylans, galactans, β-glucan, xyloglucan, chitin) and for comparisons among different foods and different samples of the same food.

The FITDOG reaction and the subsequent steps of the assay were optimized to be done in a 96-well plate format, increasing the throughput, and making it amenable to automation. In the current setup, two plates (corresponding to as many as 168 food samples and 24 calibration standards) can be reacted and prepared in parallel within 2 days and can be run on the instrument for 6 days (45 min/sample). This throughput is an improvement from conventional food composition analysis that is normally done in single vial preparations.

The FITDOG method provides more comprehensive characterization of the types and absolute amounts of different polysaccharides in food samples, compared to traditional standard enzymatic-gravimetric methods for quantitation of total starch and non-starch polysaccharides or "dietary fiber" in foods. This specificity was demonstrated using NIST SRM® 3233 (fortified breakfast cereal) where a more detailed polysaccharide composition of the sample was determined. In comparison, traditional methods only provided bulk measurements with minimal structural information. With this additional layer of information, dietary fiber composition can be further specified. In the case of apples for examples, we have shown statistically significant differences between different varieties in their arabinan, galactan, mannan, and starch contents.

This kind of resolution in carbohydrate structures can provide valuable input to other research fields, such as precision breeding in agriculture and personalized diet formulations in nutrition.

Furthermore, the FITDOG method complements the other high-throughput methods we have reported[31–34], such as monosaccharide and glycosidic linkage analyses. This suite of glycomics-based methodologies can advance research studies on food composition, including processing effects on food carbohydrates, as well as the effect of dietary carbohydrate components on the gut microbiome and their impact on health outcomes.

DATA AVAILABILITY STATEMENT

Identification and quantitation results, together with the raw mzML files were deposited in GlycoPost (ID GPST000285).

REFERENCES

(1) *Essentials of Glycobiology*, 4th edition.; Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Mohnen, D., Kinoshita, T., Packer, N. H., Prestegard, J. H., et al., Eds.; Cold Spring Harbor Laboratory Press, 2022.

(2) Seal, C. J.; Courtin, C. M.; Venema, K.; de Vries, J. Health Benefits of Whole Grain: Effects on Dietary Carbohydrate Quality, the Gut Microbiome, and Consequences of Processing. *Compr. Rev. Food Sci. Food Saf.* **2021**, *20* (3), 2742–2768. https://doi.org/10.1111/1541-4337.12728.

(3) Patnode, M. L.; Beller, Z. W.; Han, N. D.; Cheng, J.; Peters, S. L.; Terrapon, N.; Henrissat, B.; Le Gall, S.; Saulnier, L.; Hayashi, D. K.; et al. Interspecies Competition Impacts Targeted Manipulation of Human Gut Bacteria by Fiber-Derived Glycans. *Cell* **2019**, *179* (1), 59-73.e13. https://doi.org/10.1016/j.cell.2019.08.011.

(4) Koropatkin, N. M.; Cameron, E. A.; Martens, E. C. How Glycan Metabolism Shapes the Human Gut Microbiota. *Nat. Rev. Microbiol.* **2012**, *10* (5), 323–335. https://doi.org/10.1038/nrmicro2746.

(5) Tuncil, Y. E.; Thakkar, R. D.; Marcia, A. D. R.; Hamaker, B. R.; Lindemann, S. R. Divergent Short-Chain Fatty Acid Production and Succession of Colonic Microbiota Arise in Fermentation of Variously-Sized Wheat Bran Fractions. *Sci. Rep.* **2018**, *8* (1), 1–13. https://doi.org/10.1038/s41598-018-34912-8.

(6) Martens, E. C.; Lowe, E. C.; Chiang, H.; Pudlo, N. A.; Wu, M.; McNulty, N. P.; Abbott, D. W.; Henrissat, B.; Gilbert, H. J.; Bolam, D. N.; et al. Recognition and Degradation of Plant Cell Wall Polysaccharides by Two Human Gut Symbionts. *PLoS Biol.* **2011**, *9* (12). https://doi.org/10.1371/journal.pbio.1001221.

(7) Tang, W. H. W.; Kitai, T.; Hazen, S. L. Gut Microbiota in Cardiovascular Health and Disease. *Circ. Res.* **2017**, *120* (7), 1183–1196. https://doi.org/10.1161/CIRCRESAHA.117.309715.

(8) Turnbaugh, P. J.; Ley, R. E.; Mahowald, M. A.; Magrini, V.; Mardis, E. R.; Gordon, J. I. An Obesity-Associated Gut Microbiome with Increased Capacity for Energy Harvest. *Nature* **2006**, *444* (7122), 1027–1031. https://doi.org/10.1038/nature05414.

(9) 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*, 4418–4424. https://doi.org/10.1021/acs.jafc.9b00720.

(10) Li, J.; Wang, D.; Xing, X.; Cheng, T. J. R.; Liang, P. H.; Bulone, V.; Park, J. H.; Hsieh, Y. S. Y. Structural Analysis and Biological Activity of Cell Wall Polysaccharides Extracted from Panax Ginseng Marc. *Int. J. Biol. Macromol.* **2019**, *135*, 29–37. https://doi.org/10.1016/j.ijbiomac.2019.05.077.

(11) Pettolino, F. A.; Walsh, C.; Fincher, G. B.; Bacic, A. Determining the Polysaccharide Composition of Plant Cell Walls. *Nat. Protoc.* **2012**, *7* (9), 1590–1607. https://doi.org/10.1038/nprot.2012.081.

(12) An, Q.; Ye, X.; Han, Y.; Zhao, M.; Chen, S.; Liu, X.; Li, X.; Zhao, Z.; Zhang, Y.; Ouyang, K.; et al. Structure Analysis of Polysaccharides Purified from Cyclocarya Paliurus with DEAE-Cellulose and Its Antioxidant Activity in RAW264.7 Cells. *Int. J. Biol. Macromol.* **2019**. https://doi.org/10.1016/j.ijbiomac.2019.11.212.

(13) Sporck, D.; Reinoso, F. A. M.; Rencoret, J.; Gutiérrez, A.; Del Rio, J. C.; Ferraz, A.; Milagres, A. M. F. Xylan Extraction from Pretreated Sugarcane Bagasse Using Alkaline and Enzymatic Approaches. *Biotechnol. Biofuels* **2017**, *10* (1), 1–11. https://doi.org/10.1186/s13068-017-0981-z.

(14) Merkx, D. W. H.; Westphal, Y.; van Velzen, E. J. J.; Thakoer, K. V.; de Roo, N.; van Duynhoven, J. P. M. Quantification of Food Polysaccharide Mixtures by 1H NMR. *Carbohydr. Polym.* **2018**, *179*, 379–385. https://doi.org/10.1016/j.carbpol.2017.09.074.

(15) Günl, M.; Gille, S.; Pauly, M. OLIgo Mass Profiling (OLIMP) of Extracellular Polysaccharides. *J. Vis. Exp.* **2010**, No. 40, 1–5. https://doi.org/10.3791/2046.

(16) Wood, I. P.; Pearson, B. M.; Garcia-Gutierrez, E.; Havlickova, L.; He, Z.; Harper, A. L.; Bancroft, I.; Waldron, K. W. Carbohydrate Microarrays and Their Use for the Identification of Molecular Markers for Plant Cell Wall Composition. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (26), 6860–6865. https://doi.org/10.1073/pnas.1619033114.

(17) Wilkop, T.; Pattathil, S.; Ren, G.; Davis, D. J.; Bao, W.; Duan, D.; Peralta, A. G.; Domozych, D. S.; Hahn, M. G.; Drakakaki, G. A Hybrid Approach Enabling Large-Scale Glycomic Analysis of Post-Golgi Vesicles Reveals a Transport Route for Polysaccharides. *Plant Cell* **2019**, *31* (3), 627–644. https://doi.org/10.1105/tpc.18.00854.

(18) 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), 1–12. https://doi.org/10.1038/s41467-020-17778-1.

(19) Nandita, E.; Bacalzo, N. P.; Ranque, C. L.; Amicucci, M. J.; Galermo, A.; Lebrilla, C. B. Polysaccharide Identification through Oligosaccharide Fingerprinting. *Carbohydr. Polym.* **2021**, *257* (January), 117570. https://doi.org/10.1016/j.carbpol.2020.117570.

(20) Phillips, K. M.; Tarragó-Trani, M. T.; Gebhardt, S. E.; Exler, J.; Patterson, K. Y.; Haytowitz, D. B.; Pehrsson, P. R.; Holden, J. M. Stability of Vitamin C in Frozen Raw Fruit and Vegetable Homogenates. *J. Food Compos. Anal.* **2010**, *23* (3), 253–259. https://doi.org/10.1016/j.jfca.2009.08.018.

(21) Madrid Liwanag, A. J.; Ebert, B.; Verhertbruggen, Y.; Rennie, E. A.; Rautengarten, C.; Oikawa, A.; Andersen, M. C. F.; Clausen, M. H.; Scheller, H. V. Pectin Biosynthesis: GALS1 in Arabidopsis Thaliana Is a β-1,4-Galactan β-1,4-Galactosyltransferase. *Plant Cell* **2013**, *24* (12), 5024–5036. https://doi.org/10.1105/tpc.112.106625.

(22) Cankar, K.; Kortstee, A.; Toonen, M. A. J.; Wolters-Arts, M.; Houbein, R.; Mariani, C.; Ulvskov, P.; Jorgensen, B.; Schols, H. A.; Visser, R. G. F.; et al. Pectic Arabinan Side Chains Are Essential for Pollen Cell Wall Integrity during Pollen Development. *Plant Biotechnol. J.* **2014**, *12* (4), 492–502. https://doi.org/10.1111/pbi.12156.

(23) Westphal, Y.; Schols, H. A.; Voragen, A. G. J.; Gruppen, H. Introducing Porous Graphitized Carbon Liquid Chromatography with Evaporative Light Scattering and Mass Spectrometry Detection into Cell Wall Oligosaccharide Analysis. *J. Chromatogr. A* **2010**, *1217* (5), 689–695. https://doi.org/10.1016/j.chroma.2009.12.005.

(24) Mortimer, J. C.; Miles, G. P.; Brown, D. M.; Zhang, Z.; Segura, M. P.; Weimar, T.; Yu, X.; Seffen, K. A.; Stephens, E.; Turner, S. R.; et al. Absence of Branches from Xylan in Arabidopsis Gux Mutants Reveals Potential for Simplification of Lignocellulosic Biomass. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107* (40), 17409–17414. https://doi.org/10.1073/pnas.1005456107.

(25) SRM3233. Fortified Breakfast Cereal. *Natl. Inst. Stand. Technol.* **2019**, No. 281, 1.

(26) Builders, P. F.; Nnurum, A.; Mbah, C. C.; Attama, A. A.; Manek, R. The Physicochemical and Binder Properties of Starch from Persea Americana Miller (Lauraceae). *Starch/Staerke* **2010**, *62* (6), 309–320. https://doi.org/10.1002/star.200900222.

(27) Paton, D. Cellulose from Apple Tissue: Isolation, Purification and Chemical Modification. *Can. Inst. Food Sci. Technol. J.* **1974**, *7* (1), 61–64. https://doi.org/10.1016/s0315-5463(74)73849-4.

(28) Fernandes, P. A. R.; Silva, A. M. S.; Evtuguin, D. V.; Nunes, F. M.; Wessel, D. F.; Cardoso, S. M.; Coimbra, M. A. The Hydrophobic Polysaccharides of Apple Pomace. *Carbohydr. Polym.* **2019**, *223* (May), 115132. https://doi.org/10.1016/j.carbpol.2019.115132.

(29) Mankarios, A. T.; Hall, M. A.; Jarvis, M. C.; Threlfall, D. R.; Friend, J. Cell Wall Polysaccharides from Onions. *Phytochemistry* **1980**, *19* (8), 1731–1733. https://doi.org/10.1016/S0031-9422(00)83803-0.

(30) Golovchenko, V. V.; Khramova, D. S.; Ovodova, R. G.; Shashkov, A. S.; Ovodov, Y. S. Structure of Pectic Polysaccharides Isolated from Onion Allium Cepa L. Using a Simulated Gastric Medium and Their Effect on Intestinal Absorption. *Food Chem.* **2012**, *134* (4), 1813–1822. https://doi.org/10.1016/j.foodchem.2012.03.087.

(31) Couture, G.; Vo, T.-T. T.; Castillo, J. J.; Mills, D. A.; German, J. B.; Maverakis, E.; Lebrilla, C. B. Glycomic Mapping of the Maize Plant Points to Greater Utilization of the Entire Plant. *ACS Food Sci. Technol.* **2021**, *1* (11), 2117–2126. https://doi.org/10.1021/acsfoodscitech.1c00318.

(32) 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. https://doi.org/10.1021/acs.analchem.9b03101.

(33) 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. https://doi.org/10.1016/j.ijms.2018.12.009.

(34) 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. *J. Am. Soc. Mass Spectrom.* **2021**, *32* (8), 2175–2185. https://doi.org/10.1021/jasms.1c00133.

Chapter 4. A multi-glycomic platform for the analysis of food carbohydrates

ABSTRACT

Carbohydrates comprise the largest fraction of most diets and exert a profound impact on health. Components such as simple sugars and starch supply energy while indigestible components, deemed dietary fiber, reach the colon to provide food for the tens of trillions of microbes that make up the gut microbiota. The interactions between dietary carbohydrates, our gastrointestinal tracts, the gut microbiome, and host health are dictated by their structures. However, current methods for analysis of food glycans lack the sensitivity, specificity, and throughput needed to quantify and elucidate these myriad structures. This protocol describes a multi-glycomic approach to food carbohydrate analysis employing rapid-throughput liquid chromatography-tandem mass spectrometry (LC-MS/MS) methods. Quantitative analyses of monosaccharides, glycosidic linkages, and polysaccharides are performed in 96-well plate format to reduce the amount of sample and enhance throughput. Detailed stepwise processes for sample preparation, LC-MS/MS, and data analysis are provided. We illustrate the application of the protocol to a diverse set of foods as well as different apple cultivars. Furthermore, we show the utility of these methods in elucidating glycan-microbe interactions in germ-free and colonized mice. These methods provide a framework for elucidating relationships between dietary fiber, the gut microbiome, and human physiology. These structures will further guide nutritional and clinical feeding studies that enhance our understanding of the role of diet in nutrition and health.

INTRODUCTION

Carbohydrates are one of the largest components of the human diet. Humans have a limited capacity to digest these biomolecules. The *H. sapiens* genomes encode only a small number of carbohydrate-active enzymes (CAZymes) that target a limited number of glycosidic linkages found primarily in starch and simple sugars.[1] However, food carbohydrates particularly those from plants present a vast diversity of glycan structures that are not readily digested by host enzymes. Compounds containing these structures are commonly termed dietary fiber and enter the distal gut where they serve as nutrient substrates for CAZyme produced by members of the microbiota.[2-5] The output of microbial biotransformation of fiber glycans is a spectrum of metabolites that provide the host with short-term and long-term benefits.[6-10] Fiber glycan-microbiome interactions not only affect many aspects of normal physiology but also the risk of various diseases including cardiometabolic disorders, various cancers, and even the efficacy of immunotherapy for neoplasms.[11-16]

Despite their importance in common human diets, there are few analytical methods for measuring and determining the structures of carbohydrates in food.[17, 18] The most important characteristics of carbohydrates are their monosaccharide compositions and the glycosidic linkages that bind the monosaccharides together. To address the current limitations, we have developed rapid-throughput platforms based on LC-MS/MS that identify and quantify the monosaccharide compositions of carbohydrates.[19, 20] We also developed methods that quantify nearly one hundred different linkages.[21, 22] Finally, we developed a chemical method that produces unique oligosaccharides from precursor polysaccharides enabling the simultaneous identification and quantification of the polysaccharides in food.[23, 24]

DEVELOPMENT OF THE PROTOCOL

The development of these methods originated from our experience in characterizing glycans and glycoconjugates such as oligosaccharides, glycoproteins, and glycolipids in human milk as well as in various tissues.[25-31] The structural complexity of dietary carbohydrates can readily exceed those of mammalian glycans. Therefore, we first dissociate carbohydrates in various foods to determine their monosaccharide compositions. This process is achieved using rigorous acid hydrolysis, which is followed by labeling of the reducing carbon by 1-phenyl-3-methyl-5-pyrazolone (PMP). For linkage analysis, carbohydrates are first permethylated before subsequent acid hydrolysis and PMP labeling. The number and orientation of the added methyl groups allows for identification of the linkage composition of the parent structure. The resulting glycosides from both monosaccharide and linkage analysis are then analyzed separately using ultra-high performance liquid chromatography-triple quadrupole mass spectrometry (UHPLC-QqQ MS). Polysaccharide structures are delineated using FITDOG (**F**enton's **I**nitiation **T**owards **D**efined **O**ligosaccharide **G**roups): this method uses Fenton's chemistry to breakdown polysaccharides into characteristic oligosaccharides that are then reduced to their corresponding alditols before analysis by high performance liquid chromatography-quadrupole time of flight mass spectrometry (HPLC-qToF). The observed oligosaccharide signals then provide identification and quantification of the parent polysaccharides. **Figure 4.1** provides an overview of the three methods that comprise the multi-glycomic analysis.

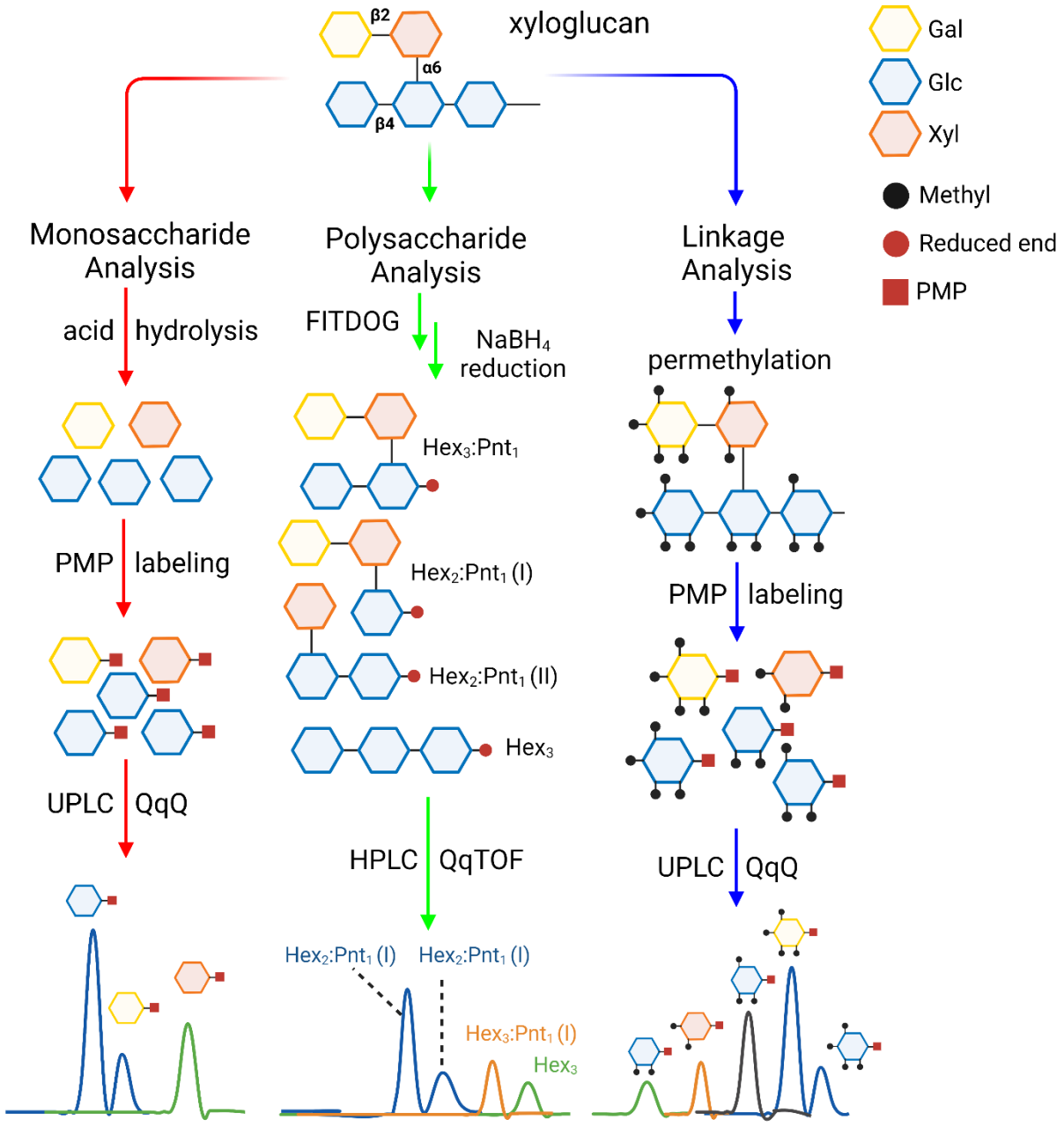


Figure 4.1. A representation of the multi-glycomic workflow. Food carbohydrates are quantified and structurally elucidated by monosaccharide compositional analysis (left panel), polysaccharide analysis (center), and glycosidic linkage analysis (right) using rapid throughput chemical and LC-MS/MS methods.

The methods employ 96-well plate formats, and UHPLC is utilized, when possible, to increase throughput and decrease sample run time. Additionally, the entire protocol can be

performed with just a few milligrams of sample allowing the determination of monosaccharide composition, glycosidic linkages, and amounts of polysaccharide with minimal amounts of material (sub-microgram, sub-microgram, and sub-milligram, respectively).

COMPARISONS TO ALTERNATIVE METHODS

The most commonly employed method for the characterization and quantification of carbohydrates in plants and food by monosaccharide and linkage analysis uses gas chromatography-mass spectrometry (GC-MS).[32] GC-MS instruments are significantly cheaper than LC-MS/MS instruments and offer adequate chromatographic resolution. The ability to use electron impact (EI) ionization also makes identification of monosaccharides and their permethylated derivatives by mass spectral matching an attractive feature. For monosaccharide analysis, carbohydrates must be hydrolyzed, acetylated, and reduced to their corresponding alditol acetates (AAs) prior to injection.[32, 33] For linkage analysis, an initial permethylation is needed to obtain corresponding permethylated alditol acetates (PMAAs).[32, 34, 35] However, this analytical approach has significant limitations; GC-MS instrument run times are typically longer for monosaccharide analysis (20 min vs. 5 min) and for linkage analysis (60 min vs. 16 min). It requires pre-derivatization of the liberated monosaccharides to allow GC analysis. Long run times and the need for samples to be in highly volatile solvents for GC analysis makes the adaptation of these methods to 96-well plates difficult. Thus, samples are prepared one vial at a time severely limiting sample throughput. Additionally, because only a single quadrupole instrument is typically used, GC-MS chromatograms are prone to high noise and relatively low sensitivity (milligram for GC-MS vs. picogram for LC-MS monosaccharide analysis) especially if further upstream purifications are not performed.[19, 32] Lastly, the coverage of glycosidic

linkages by GC-MS is also limited compared to the current LC-MS/MS approach. Most importantly, polysaccharide analysis analogous to FITDOG is likely not possible with GC-MS.

High-performance anion-exchange chromatography coupled to pulsed amperometric detection (HPAEC-PAD) is also commonly employed for monosaccharide and oligosaccharide analysis. HPAEC-PAD is particularly attractive for carbohydrate analysis because it is highly selective, sensitive, and does not require derivatization prior to analysis.[36] However, isomer separation can be challenging and often requires long method run times.[36, 37] Furthermore, linkage analysis by HPAEC appears not to be possible due to the large number of potential isomers and the lack of structural information in the method.

The LC-MS/MS methods presented here directly address many of the disadvantages of GC-MS and HPAEC-PAD. Namely, the chromatographic separation time of PMP-labeled glycosides on a C18 stationary phase is drastically reduced relative to AAs and PMAAs in GC-MS analysis and to underivatized carbohydrates in HPAEC-PAD. For example, the monosaccharide method described here separates 14 monosaccharides in 4.6 min while a similar separation using GC-MS may require upwards of 20 min (8 h vs. 48 h for 96 samples).[19, 32, 33, 38] Because most of the sample preparation for this protocol is in aqueous solvent and significantly less prone to evaporation, it is also more adaptable to a 96-well plate format (unlike GC-MS). Furthermore, the use of MS/MS improves the sensitivity and the specificity, particularly in monosaccharide and linkage analyses , thereby requiring fewer preparation steps than is typical for GC-MS. Additionally, the LC-MS/MS approach expands greatly the limited linear and dynamic ranges of the GC and HPAEC approaches, offering ranges of up to 6 orders of magnitude relative to the 2 to 3 orders of magnitude provided by the latter methods.[19]

Polysaccharide analysis by FITDOG is a very recent method that can be used for direct identification and quantification of polysaccharides.[23, 39, 40] The method is highly specific because characteristic oligosaccharides are produced from each polysaccharide, thus differentiating each polymer. Direct identification of polysaccharides in food has historically required extensive sample preparation and extraction techniques to isolate individual polysaccharides for GC-MS, HPAEC-PAD, and/or nuclear magnetic resonance (NMR) analysis.[32, 36, 41-43] Quantification could only be performed gravimetrically. FITDOG analysis can be performed with high throughput on mixtures of polysaccharides such as those present in food or even in feces for clinical trials.

The approach used for polysaccharide analysis is similar to other 'bottom-up' approaches used in genomics and proteomics where the polysaccharides are first broken down into smaller oligosaccharide fragments. The matching of the resulting oligosaccharide compounds to their parent polysaccharide structure is based on an oligosaccharide fingerprinting library obtained by reacting commercially-available polysaccharide standards. An ethanol precipitation step prior to polysaccharide analysis helps to ensure that endogenous oligosaccharides present in the samples are not attributed to polysaccharide structures. We have used the non-enzymatic and reproducible reaction to depolymerize common plant polysaccharides using radical reaction.[23, 24, 39] A catalytic amount of Fe^{3+} and an excess of H_2O_2 are added to the reaction mixture to produce reactive oxygen radicals that facilitate the cleavage of glycosidic bonds in the polysaccharides. Key to this process was the optimization to minimize and eliminate higher order reactions that further degrade the oligosaccharide products. The resulting oligosaccharides from the depolymerization reaction are then reduced using $NaBH_4$ to prevent anomer separation that we observe on some oligosaccharides separated on a PGC column. The oligosaccharide reduction

helps further to reduce the complexity in matching retention times (RT) and accurate masses of the oligosaccharides to the fingerprinting library. Finally, reduced oligosaccharides are enriched and cleaned using C18 and PGC SPE. Oligosaccharide profiles are then obtained using HPLC-qToF in data-dependent mode.

OVERVIEW OF THE PROTOCOL

The overall workflow is provided in **Figure 4.1.** The more detailed descriptions of the analyses are provided in **Figure 4.2**. Food, fecal, or plant tissue samples are first lyophilized before homogenization into a powder using a bead mill or a mortar and pestle. For hard and rigid samples, a coffee grinder may be used first to homogenize. Milligram quantities of the homogenized sample are then weighed into 1.5 mL screw-cap tubes and further homogenized with the addition of water, bead milling, and incubation at 100 °C. Ethanol (EtOH) precipitation may also be used at this step to separate high and low molecular-weight carbohydrates if separate analyses are needed for soluble and insoluble fractions. Each fraction may then be analyzed separately. If no EtOH precipitation is performed, the resulting suspensions are directly aliquoted into individual 96-well plates, one for monosaccharide, another for glycosidic linkage, and yet another for polysaccharide (FITDOG) analysis. If an EtOH precipitation is performed, the supernatant is removed, and the pellet is washed and subsequently homogenized according to the aforementioned steps. For quantitative monosaccharide compositional analysis, sample aliquots are hydrolyzed with trifluoroacetic acid (TFA), derivatized with 1-phenyl-3-methyl-5-pyrazolone (PMP), extracted with chloroform ($CHCl_3$) to remove excess PMP, separated using UHPLC on a C18 column, and analyzed on a QqQ MS instrument operated in dynamic multiple reaction monitoring (dMRM) mode. Absolute quantification is achieved using an external calibration

curve. However, isotopically labeled internal standards may also be employed. Glycosidic linkage analysis employs the same steps with the addition of a permethylation step prior to acid hydrolysis. Linkages are assigned using an in-house library of retention times and MRM transitions. Polysaccharide analysis is performed by using the FITDOG reaction in which polysaccharides in an EtOH precipitated sample aliquot are oxidatively cleaved using Fenton's chemistry into characteristic oligosaccharide fragments. The resulting oligosaccharides are then reduced with sodium borohydride ($NaBH_4$) and subjected to a solid phase extraction (SPE) cleanup with both C18 and porous graphitized carbon (PGC) before analysis on an HPLC-qToF instrument equipped with a PGC column. In summary, this workflow provides a means to obtain three levels of information on the carbohydrates in food, feces, and plant tissues.

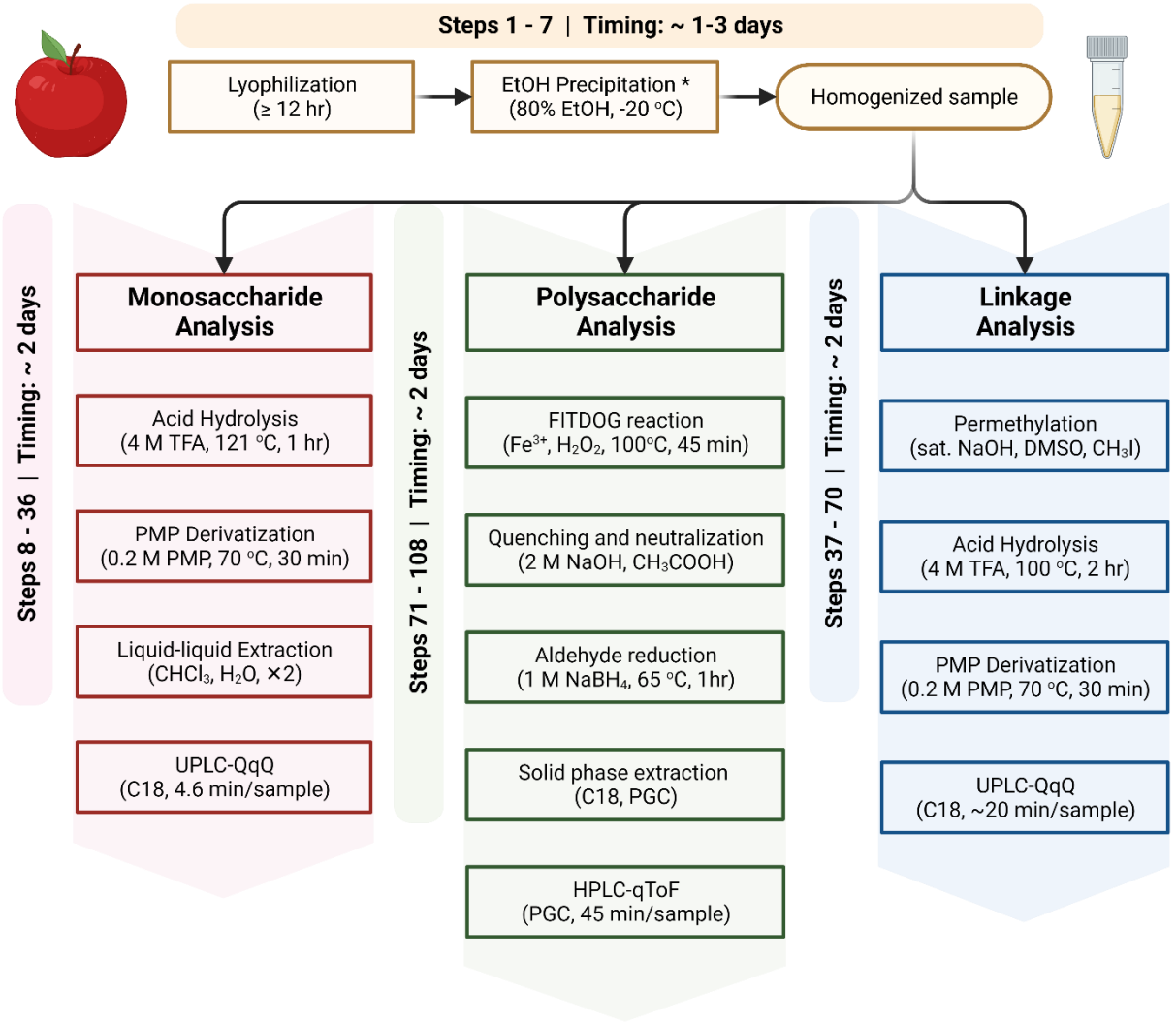


Figure 4.2. Summary of the workflow and timing for comprehensive analysis of carbohydrates in food and feces. The major steps for each of the three analyses are summarized. Monosaccharide analysis is shown in the left flowchart, polysaccharide FITDOG analysis in the center, and linkage analysis in the right.

Sample preparation and homogenization (Steps 1-7)

Because the protocol uses small aliquot volumes and masses, it is paramount that samples be thoroughly homogenized prior to entering the analytical workflow. We have found that sampling from dry material improves precision and sensitivity by normalizing moisture content

to a minimum, thereby concentrating the carbohydrates in the sample. Lyophilization (or "freeze-drying") has worked very well for this purpose. Samples should be flash-frozen at –80 °C for a minimum of 3 h to prepare them for lyophilization. Once frozen, samples can be dried; the length of time required to reach a minimum moisture content depends on several factors including the freeze-dryer being used, the original moisture content of the sample, and the sample matrix itself. In general, we have found that about three days ensures adequate drying of all sample types. If moisture content data is desired, weighing the sample vessel before and after drying is a necessary step. Very hygroscopic food samples will retain moisture even after a few days in the lyophilizer. If total moisture content and absolutely dry basis values are needed, a separate method (e.g. Karl-Fischer titration) can be used. However, this additional step is not presented in this protocol.

Once samples are dried, they must be homogenized into a fine powder to allow precise aliquoting. Additional instruments may be needed for homogenization. For hard samples such as seeds, a coffee grinder is highly recommended. For hygroscopic and high-fat samples such as dried fruit and nuts, we have found that flash-freezing with liquid nitrogen and grinding in a mortar and pestle is best. Many samples, such as vegetables, legumes, meat, grains, and leaf tissues (as well as feces) are readily homogenized by bead milling alone using stainless-steel beads.

After drying, samples are weighed so that solutions can be produced for analysis. Samples should be weighed into screw-cap tubes that are compatible with the bead mill homogenizer for the subsequent homogenization steps. We typically weigh aliquots of about 10 mg and add 1 mL of water to create a solution of 10 mg/mL for ease of measurement, throughput, and storage. However, if samples are limited, as little as 1 mg may be used. The

limitation is not with the sensitivity of the methods, but rather with the accuracy and precision of the analytical balance and the homogeneity of the starting sample. If the bulk sample is to be analyzed, then 1 mL of water and stainless-steel beads are added to the sample tube. Aqueous solutions are then subjected to bead-mill homogenization, incubation at 100 °C for 1 h, followed by an additional round of bead-milling.

Polysaccharides may be separated from the smaller components (mono-, di-, and oligosaccharides) and analyzed separately by first adding 80% EtOH to the weighed unprocessed sample. The samples are vortexed and centrifuged. The soluble fractions are removed by pipette and analyzed separately. However, analysis of the soluble fraction is not included in this protocol; we instead focus on the insoluble polysaccharide-containing fraction. After the supernatant is removed, the pellet is washed with two volumes of 80% EtOH and dried with vacuum centrifugation. Once dried, the Eppendorf tube containing the pellet is subjected to homogenization with the addition of water and stainless-steel beads. After homogenization, the Eppendorf tube is heated to 100 °C for 1 h, followed by additional bead-milling. Sample solutions may be stored at –20 °C or –80 °C prior to analysis. If the samples are to be stored, they must be thawed and homogenized again via bead-milling prior to analysis.

Monosaccharide composition: hydrolysis, derivatization, extraction, and UHPLC-QqQ MS analysis (Steps 8-36)

The homogenized solutions are then aliquoted either to a polypropylene 96-well plate or to individual 1.5 mL screw-cap Eppendorf vials. The homogenized solutions appear as suspensions and must be well-vortexed prior to pipetting. For the hydrolysis, water and TFA are

added to the suspension and the mixture is heated to 121 ℃ for 1 h. This condition provides quantitative hydrolysis for most polysaccharide components while minimizing degradation of the liberated monosaccharides. However, some optimization may be required for other samples such as those containing fructans, since fructose degrades at these conditions. Thus, fructose may be underrepresented under these conditions. If a more accurate analysis of fructose is required, gentler hydrolysis conditions are recommended (100 ℃ for 1 h). For large-scale analysis, a 96-well plate that can withstand the high temperature, a plate lid, and a clamp to seal the plate are necessary. The 96-well plate and plate lids are commercially available, and the clamp can be machined as shown in **Figure 4.3**. Once the hydrolysis is complete, ice-cold water is added.

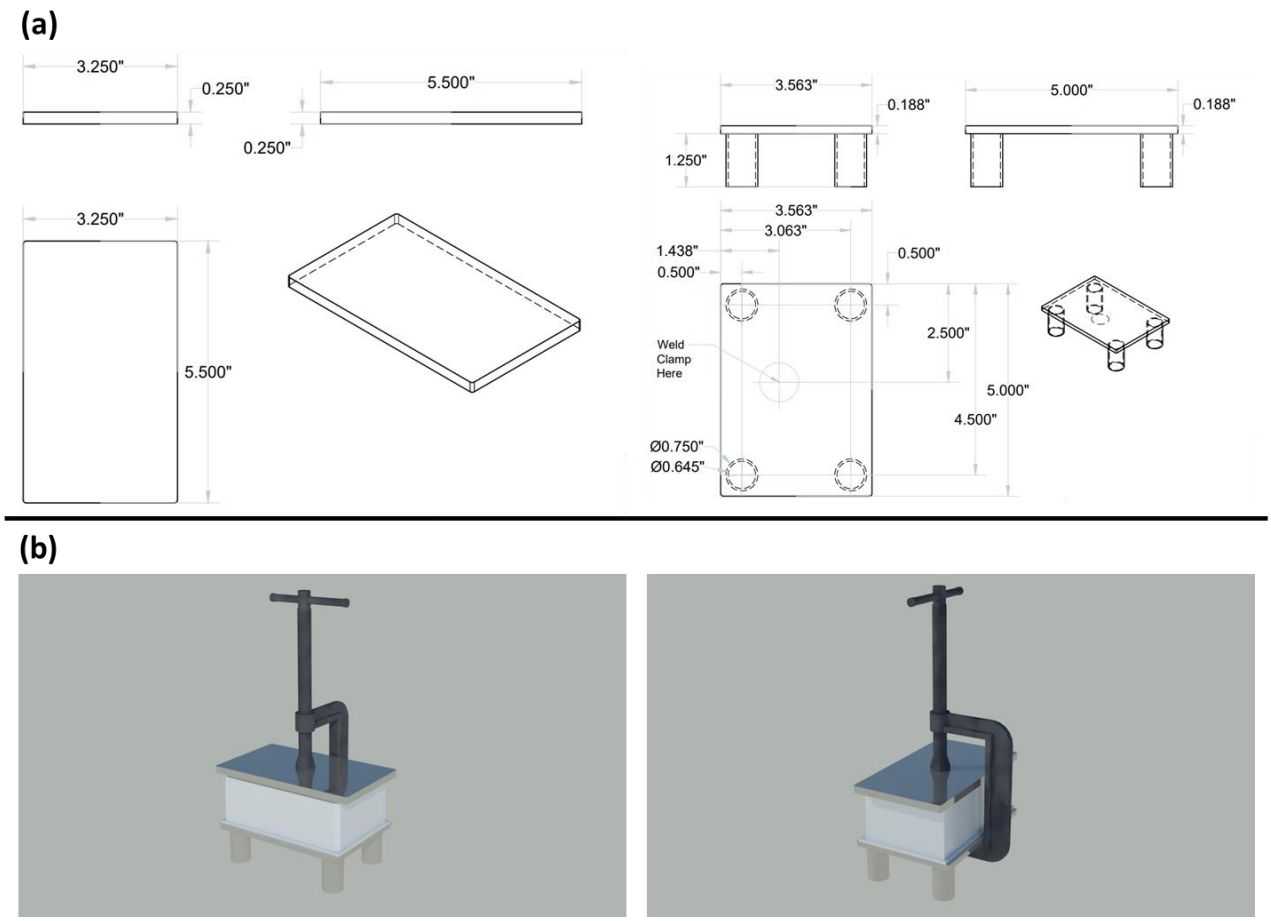


Figure 4.3. Schematic diagram of the stainless-steel plate clamps used for monosaccharide and linkage analysis. (a) Schematic of top and bottom plates with dimensions compatible with the 96-well plates described in the protocol. (b) Illustrations of finished clamps with C-clamps included. C-clamp is welded to the bottom of the bottom plate.

After hydrolysis, the resulting monosaccharides are labeled with PMP. Additionally, a pooled set of external standards, containing the 14 most common monosaccharides in food (D-fructose, D-allose, D-mannose, D-glucose, D-galactose, L-rhamnose, L-fucose, D-ribose, D-xylose, L-arabinose, D-glucuronic acid, D-galacturonic acid, N-acetyl-D-glucosamine, N-acetyl-D-galactosamine), are prepared to produce an external calibration curve. Aliquots of the hydrolysate are transferred to another 96-well plate where a methanolic PMP and ammonia solution are added for labeling. The reaction is carried out at 70 ˚C, thus the plate must be

clamped. Once the derivatization is complete, the samples are dried by vacuum centrifugation. A programmable vacuum centrifuge is recommended to prevent solvent bumping. The dried and labeled monosaccharides are then extracted with $CHCl_3$ to remove excess PMP. The aqueous layer containing the analytes is removed and analyzed with a UHPLC-QqQ MS equipped with a C18 column. Peak areas for each monosaccharide obtained from samples are compared to the external calibration curve for quantitation. A representative chromatogram of a pooled monosaccharide standard solution is depicted in **Figure 4.4A**.

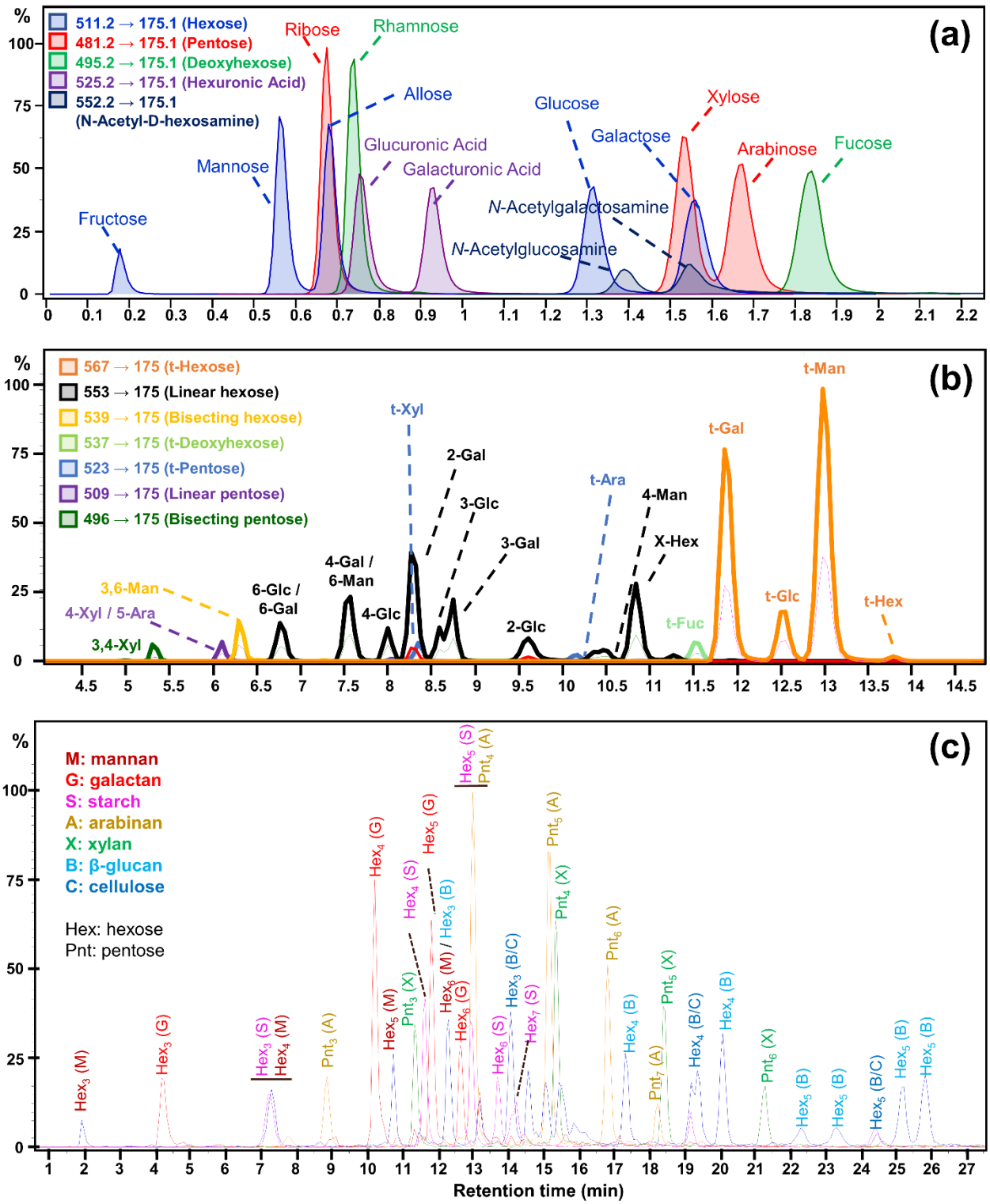


Figure 4.4. dMRM transitions and extracted ion chromatograms (EICs) of pooled monosaccharides, linkages, and polysaccharides. (a) UHPLC-dMRM quantifier ion transitions for the 14 monosaccharides monitored by the method. (b) UHPLC-dMRM quantifier transitions for a pool of oligosaccharide standards containing the preponderant linkages found in food. (c)

EICs of oligosaccharides generated by FITDOG from a pool containing common food polysaccharides.

Glycosidic linkage compositions: permethylation, hydrolysis, derivatization, extraction, and UHPLC-QqQ MS analysis (Steps 37-70)

The procedure for linkage analysis is similar to the monosaccharide analysis with the addition of a permethylation step prior to acid hydrolysis and labeling. Once samples are aliquoted into a 96-well plate, they are dried by vacuum centrifugation. Saturated NaOH is added followed by DMSO to solubilize and activate the carbohydrates. Iodomethane (CH_3I) is then introduced to permethylate the samples, and the reaction is quenched by the addition of water. Dichloromethane (DCM) is added to solubilize the permethylated products. The DCM layer is extracted with five volumes of water to remove salts and DMSO. The organic layer is dried by vacuum centrifugation before being subjected to acid hydrolysis with 4 M TFA (100 ˚C for 2 h). After hydrolysis, samples are dried again before PMP derivatization. The resulting permethylated and labeled glycosides are dried a final time before reconstitution in 70% MeOH prior to UHPLC-QqQ MS analysis. An in-house library containing the MRM transitions and retention times of the most commonly observed unique linkages in food is used to identify the observed linkages (**Table 4.1**). While this library consists of the preponderant 47 linkages observed in food, up to 96 can be monitored in each sample using an expanded library.[21] An example chromatogram from a carbohydrate standard pool is shown in **Figure 4.4B**.

Table 4.1. MRM transition list for the glycosidic linkage analysis using UPLC-QqQ.

Name	TS	Transition	Scan	Type	Precursor Ion	Product Ion	RT	Ion Polarity
2,4,6-Glucose	2	525.2 -> 175.1	MRM	Target	525.2	175.1	4.135	Positive
2,3,6-Glucose	2	525.2 -> 175.1	MRM	Target	525.2	175.1	4.263	Positive
2-Mannose	2	553.6 -> 175.1	MRM	Target	553.2	175.1	4.436	Positive
2,4,6-Galactose	2	525.2 -> 175.1	MRM	Target	525.2	175.1	4.455	Positive
3,4,6-Glucose	2	525.2 -> 175.1	MRM	Target	525.2	175.1	4.647	Positive
2-Rhamnose	2	523.2 -> 175.1	MRM	Target	523.2	175.1	4.716	Positive
2,4-Xylose	2	495.2 -> 175.1	MRM	Target	495.2	175.1	4.918	Positive
3,4,6-Galactose	2	525.2 -> 175.1	MRM	Target	525.2	175.1	5.031	Positive
2,3-Arabinose	2	495.2 -> 175.1	MRM	Target	495.2	175.1	5.238	Positive
4,6-Glucose	2	539.2 -> 175.1	MRM	Target	539.2	175.1	5.279	Positive
3,4-*p*-Xylose/3,5-Arabinose	2	495.2 -> 175.1	MRM	Target	495.2	175.1	5.431	Positive
3,6-Galactose	2	539.2 -> 175.1	MRM	Target	539.2	175.1	5.471	Positive
3,4,6-Mannose	2	525.2 -> 175.1	MRM	Target	525.2	175.1	5.543	Positive
2,5-Arabinose	2	495.2 -> 175.1	MRM	Target	495.2	175.1	5.751	Positive
2,4-Glucose	2	539.2 -> 175.1	MRM	Target	539.2	175.1	6.111	Positive
4-*p*-Xylose	2	509.2 -> 175.1	MRM	Target	509.2	175.1	6.258	Positive
5-*f*-Arabinose	2	509.2 -> 175.1	MRM	Target	509.2	175.1	6.386	Positive
3,4-Galactose	2	539.2 -> 175.1	MRM	Target	539.2	175.1	6.495	Positive
3,6-Mannose	2	539.2 -> 175.1	MRM	Target	539.2	175.1	6.559	Positive
3,4-Glucose	2	539.2 -> 175.1	MRM	Target	539.2	175.1	6.815	Positive
6-Glucose	2	553.6 -> 175.1	MRM	Target	553.2	175.1	6.933	Positive
2-Xylose	2	509.2 -> 175.1	MRM	Target	509.2	175.1	7.026	Positive
4,6-Galactose	2	539.2 -> 175.1	MRM	Target	539.2	175.1	7.072	Positive
6-Galactose	2	553.6 -> 175.1	MRM	Target	553.2	175.1	7.125	Positive
3-Arabinose	2	509.2 -> 175.1	MRM	Target	509.2	175.1	7.538	Positive
4,6-Mannose	2	539.2 -> 175.1	MRM	Target	539.2	175.1	7.648	Positive
4-Galactose	2	553.6 -> 175.1	MRM	Target	553.2	175.1	7.766	Positive
6-Mannose	2	553.6 -> 175.1	MRM	Target	553.2	175.1	8.086	Positive
4-Glucose	2	553.6 -> 175.1	MRM	Target	553.2	175.1	8.406	Positive

2-*f*-Arabinose	2	509.2 -> 175.1	MRM	Target	509.2	175.1	8.755	Positive
2-Galactose	2	553.6 -> 175.1	MRM	Target	553.2	175.1	8.79	Positive
t-*p*-Xylose	2	523.2 -> 175.1	MRM	Target	523.2	175.1	8.942	Positive
t-Glucuronic acid	2	581.2 -> 175.1	MRM	Target	581.2	175.1	8.965	Positive
3-Glucose/3-Galactose	2	553.6 -> 175.1	MRM	Target	553.2	175.1	9.174	Positive
t-*p*-Arabinose	2	523.2 -> 175.1	MRM	Target	523.2	175.1	9.744	Positive
2-Glucose	2	553.6 -> 175.1	MRM	Target	553.6	175.1	10.199	Positive
t-*f*-Arabinose	2	523.2 -> 175.1	MRM	Target	523.2	175.1	10.67	Positive
t-Galacturonic acid	2	581.2 -> 175.1	MRM	Target	581.2	175.1	10.694	Positive
3-Mannose	2	553.6 -> 175.1	MRM	Target	553.2	175.1	10.839	Positive
4-Mannose	2	553.6 -> 175.1	MRM	Target	553.2	175.1	11.095	Positive
x-Hexose	2	553.6 -> 175.1	MRM	Target	553.2	175.1	11.799	Positive
t-Fucose	2	537.2 -> 175.1	MRM	Target	537.2	175.1	12.263	Positive
t-Galactose	2	567.2 -> 175.1	MRM	Target	567.2	175.1	12.559	Positive
t-Rhamnose	2	537.2 -> 175.1	MRM	Target	537.2	175.1	12.903	Positive
t-Glucose	2	567.2 -> 175.1	MRM	Target	567.2	175.1	13.391	Positive
t-Mannose	2	567.2 -> 175.1	MRM	Target	567.2	175.1	13.904	Positive
t-Hexose	2	567.2 -> 175.1	MRM	Target	567.2	175.1	14.352	Positive

Polysaccharide composition: FITDOG, reduction, purification, and HPLC-qToF analysis (Steps 71-108)

Prior to FITDOG analysis, we recommend that samples be precipitated with EtOH, using the procedure previously detailed, to remove existing mono-, di, and oligosaccharides. Homogenized samples are then aliquoted to a 96-well plate where solutions of iron (III) sulfate and hydrogen peroxide are added to produce reactive oxygen radicals and facilitate glycosidic bond cleavage. The reaction is quenched by the addition of NaOH and subsequently neutralized with acetic acid. To avoid volume limitations, aliquots of the depolymerized samples are transferred to another 96-well plate where they are reduced with $NaBH_4$. The resulting reduced oligosaccharides are enriched using C18 and PGC SPE.

Samples are analyzed via an HPLC-qToF equipped with a PGC column. Oligosaccharide peaks are annotated based on accurate precursor mass and their observed tandem mass spectra. Oligosaccharides are then assigned to their parent polysaccharide structures by matching RT values and compositions to the fingerprint library. We recommend that several standards be run in parallel with the samples to ensure proper RT matching. Starch and cellulose are relatively inexpensive and commercially available and thus can be used for quality control purposes. Peaks from these standards are utilized to account for RT shifts when matching compounds to the fingerprint library. An example of the resulting LC-MS chromatogram is provided in **Figure 4.4C**.

LIMITATIONS OF THE PROTOCOL

One limitation of this protocol is the requirement of LC-MS/MS instruments, which are generally large and expensive. Current standard carbohydrate methods employ GC-MS and HPAEC-PAD, which are also large albeit somewhat cheaper than LC-MS instruments. Furthermore, at most research and academic institutions, LC-MS instruments have become more available and even essential in central core facilities.

Another limitation of the current iteration of the protocol relates to coverage of acidic polysaccharides such as galacturonan and rhamnogalacturonan. These structures are still represented by the current protocol, but glycosidic bonds adjacent to uronic acids are recalcitrant to acid hydrolysis. These polysaccharides require enzymatic digestion or reduction of the carboxylic acid prior to hydrolysis for more quantitative coverage. One or both steps may be added to future protocols if better coverage of these polysaccharides is required.

Finally, data analysis can be cumbersome and time-consuming, particularly for linkage and FITDOG analyses. Due to the nature of glycan structures, there can be a significant number of isomeric species co-eluting throughout the LC-MS run. It then becomes difficult for these compounds to be quantified properly. Automated software is used to annotate and quantitate the compound; however, manual confirmation is often necessary when overlapping peaks are present. Recent advances in machine learning for automated peak integration may soon solve this limitation.[44]

FUTURE APPLICATIONS

Carbohydrate structures and abundances in the foods we consume are still poorly defined; the methods in this protocol can be applied to any food for quantification and structural characterization of their carbohydrates. They can also be incorporated into existing fiber analysis protocols like AOAC 991.43 or 2017.16 to determine the compositions of isolated insoluble and soluble dietary fiber fractions.

This protocol can also be used to explore the utilization of fiber by gut microbes both *in vitro* and *in vivo*. A current major challenge is connecting measured carbohydrate-active enzyme gene and transcript abundances from metagenomic and metatranscriptomic experiments to an explicit functional outcome. By applying the described protocol to the feces of study participants or from small-scale bioreactors, fecal glycan abundances and structures can be determined thus providing mechanistic insight towards food composition, the microbiome, and CAZyme function.[13, 14, 45, 46] The data can also inform differences in the responses of study participants to dietary interventions.[13, 14]

MATERIALS

Reagents

- Acetonitrile (ACN, Honeywell, cat. no. 34967)

! CAUTION Acetonitrile is a flammable liquid and vapor. Wear personal protective equipment and use only in a chemical fume hood.

- Acetic acid (Glacial, Supelco; MilliporeSigma, cat. no. AX0073)

! CAUTION Acetic acid is a flammable liquid and causes severe skin irritation and eye damage. Wear personal protective equipment when handling.

- Ammonium hydroxide solution (NH_4OH, 28-30%, NH3 basis, MilliporeSigma, cat. no. 221228)

! CAUTION Ammonium hydroxide solution is corrosive and causes severe skin burns and eye damage. Wear personal protective equipment and use only in a chemical fume hood.

- Ammonium acetate (NH_4Ac, 99.999% trace metals basis, MilliporeSigma, cat. no. 372331)

! CAUTION Ammonium acetate is a combustible solid. Wear personal protective equipment when handling.

- Chloroform ($CHCl_3$, MilliporeSigma, cat. no. 34854)

! CAUTION Chloroform is a flammable liquid and vapor, acutely toxic, and causes eye and skin irritation. Wear personal protective equipment when handling and use only in a chemical fume hood.

- Ethanol (EtOH, MilliporeSigma, cat. No. E7023)

! CAUTION Ethanol is a flammable liquid and vapor, acutely toxic, and causes eye and skin irritation. Wear personal protective equipment when handling and use only in a chemical fume hood.

- Formic Acid (FA, Optima™ LC/MS Grade; Fisher Chemical™, cat. no. A117)

! CAUTION Formic acid is a flammable liquid and vapor and causes severe skin burns and eye damage. Wear personal protective equipment when handling and use only in a chemical fume hood.

- Hydrogen peroxide solution (H_2O_2, 30 wt. % in H_2O, MilliporeSigma, cat.no. 216763)

! CAUTION Hydrogen peroxide is a strong oxidizer and causes severe skin burns and eye damage. Wear personal protective equipment and use only in a chemical fume hood.

- Methanol (MeOH, MilliporeSigma, cat. no. 34860)

! CAUTION Methanol is a flammable liquid and vapor. It is toxic by inhalation and contact with skin and eyes. Wear personal protective equipment and use only in a chemical fume hood.

- Iron (III) sulfate pentahydrate ($Fe_2(SO_4)_3 \cdot 5H_2O$, 97%, Thermo Scientific, cat.no. AC345235000)

! CAUTION Iron (III) sulfate pentahydrate causes skin irritation and eye damage. Wear personal protective equipment when handling.

- Sodium acetate ($NaCH_3CO_2$, MilliporeSigma, cat. no. 79714)
- Sodium hydroxide (NaOH, 99.99% trace metals basis; MilliporeSigma, cat. no. 306576)

! CAUTION Sodium hydroxide causes severe skin burns and eye damage. Wear personal protective equipment when handling.

- Sodium hydroxide (NaOH, ACS reagent ≥97.0%; MilliporeSigma, cat. 795429)

! CAUTION Sodium hydroxide causes severe skin burns and eye damage. Wear personal protective equipment when handling.

- Sodium borohydride ($NaBH_4$, MilliporeSigma, cat.no. 806373)

! CAUTION Sodium borohydride causes skin corrosion, eye damage, and reproductive toxicity. Contact with water also releases flammable gas. Wear personal protective equipment when handling.

- Trifluoroacetic acid (TFA, Optima™ LC/MS Grade; Fisher Chemical™, cat. no. A116)

! CAUTION Trifluoroacetic acid causes severe skin burns and eye damage. Wear personal protective equipment when handling. Use only in a chemical fume hood.

- 3-Methyl-1-phenyl-2-pyrazoline-5-one (PMP, MilliporeSigma, cat.no. M70800)

! CAUTION PMP is toxic and causes eye irritation. Wear personal protective equipment when handling.

Monosaccharide Analysis Standards

- D-(+)-Glucose (MilliporeSigma, cat. no. G8270)
- D-(+)-Galactose (MilliporeSigma, cat. no. G0750)
- D-(-)-Fructose (MilliporeSigma, cat. no. F2793)
- D-(+)-Mannose (MilliporeSigma, cat. no. 92683)
- D-(+)-Allose (MilliporeSigma, cat. no. 285005)
- D-(-)-Ribose (MilliporeSigma, cat. no. R7500)
- D-(+)-Xylose (MilliporeSigma, cat. no. X1500)
- L-(+)-Arabinose (MilliporeSigma, cat. no. A3256)
- L-(+)-Rhamnose monohydrate (MilliporeSigma, cat. no. 41651)
- L-(-)-Fucose (MilliporeSigma, cat. no. 93183)

- D-(+)-Glucuronic acid (MilliporeSigma, cat. no. G5269)
- D-(+)-Galacturonic acid (MilliporeSigma, cat. no. 92478)
- N-Acetyl-D-Glucosamine (MilliporeSigma, cat. no. A8625)
- N-Acetyl-D-Galactosamine (MilliporeSigma, cat. no. A2795)

Linkage Analysis Standards

- 2-O-(α-D-Mannopyranosyl)-D-mannopyranose (Biosynth, cat. no. OM05906)
- 1,3-α-1,6-α-D-Mannotriose (Biosynth, cat. no. OM05762)
- 1,4-β-D-Mannotriose (Biosynth, cat. no. OM31999)
- 1,4-β-D-Xylobiose (Biosynth, cat. no. OX05190)
- 1,5-α-L-Arabinotriose (Biosynth, cat. no. OA32462)
- 3^3-α-L-Arabinofuranosyl-xylotetraose (Megazyme, cat. no. O-XA3XX)
- Isomaltotriose (Biosynth, cat. no. OI05352)
- Maltohexaose (Biosynth, cat. no. OM06869)
- Nigerose (Biosynth, cat. no. ON06975)
- Sophorose (Biosynth, cat. no. OS06893)
- Amylopectin (Biosynth, cat. no. YA39745)
- 3-O-(β-D-Galactopyranosyl)-D-galactopyranose (Biosynth, cat. no. OG10186)
- 4-O-(β-D-Galactopyranosyl)-D-galactopyranose (Biosynth, cat. no. OG04727)
- Lactose (Biosynth, cat. no. OL04771)
- 2'-Fucosyllactose (Biosynth, cat. no. OF06739)
- Sophorose monohydrate (Biosynth, cat. no. OS06893)

Polysaccharide (FITDOG) Analysis Standards

- Starch (corn, analytical grade, MilliporeSigma, cat. no. S5296)
- Chitin (shrimp shells, MilliporeSigma, cat. no. C9752)
- Cellulose (microcrystalline powder, extra pure, ACROS Organics)
- Linear arabinan (sugar beet pulp, Megazyme, cat. no. P-LARB)
- Mannan (ivory nut seeds, Megazyme, cat. no. P-MANIV)
- Galactan (potato, Megazyme, cat. no. P-GALPOT)
- Xylan (beechwood, Megazyme, cat. no. P-XYLNBE)
- Xyloglucan (tamarind seeds, Megazyme, cat. no. P-XYGLN)
- β-glucan (barley, Megazyme, cat. no. P-BGBM)

EQUIPMENT

- Falcon centrifuge tubes (50 mL; Corning, cat.no. 352070)
- Falcon centrifuge tubes (15 mL; Corning, cat.no. 352196)
- Freeze Dryer (SP Scientific, cat.no. BTP-3ESE0W)
- Pipettes (Gilson, cat. no. FA10003M, FA10005M, FA10006M)
- Multichannel Pipettes (USA Scientific, cat. no. 7112-0510, 7112-1100, 7112-3000; Eppendorf, cat. no. 3125000222)
- Bead Mill Homogenizer (OMNI International, cat.no. 19-042E)
- Analytical Balance (METTLER TOLEDO, cat.no. XS105)
- 30 mL Tubes (OMNI International, cat.no. 19-6635)
- 96-Well Polypropylene DeepWell Plate (Thermo Scientific, cat.no. 95040452)

- Sealing Mats for 96-Well Plate (Thermo Scientific, cat.no. AB-0675)
- Plate shaker (Scientific Industries, model no. SI-4000)
- Centrifugal Vacuum Concentrator (SP Scientific, model no. QUC-12060-C00)
- Pipette tips (USA Scientific, cat. no. 1111-3800, 1110-9800, 1112-1820)
- 1.5 mL screw cap tube (Sarstedt, cat.no. 72.692.005)
- C18 SPE PLATE (40 µL filter plate, C-18, Glysci, cat. FNSC18)
- Graphitized carbon (PGC) SPE plate (40 µL filter plate, Carbon (Hypercarb), Glysci, cat. FNSCAR)
- Poroshell HPH C18 UHPLC Column (1.9 µm, 2.1 × 50 mm, Agilent Technologies, cat. no. 699675-702)
- Poroshell HPH C18 UHPLC guard cartridges (1.9 µm, 2.1 × 5 mm, Agilent Technologies, cat. no. 821725-945)
- ZORBAX RRHD Eclipse Plus C18 UHPLC Column (1.8 µm, 2.1 × 150 mm, Agilent Technologies, cat. no. 959759-902)
- ZORBAX RRHD Eclipse Plus C18 UHPLC guard cartridges (1.8 µm, 2.1 × 5 mm, Agilent Technologies, cat. no. 821725-901)
- Analytical PGC (Hypercarb™) HPLC Column (5 µm, 1 × 150 mm, Thermo Scientific, cat. no. 35005-151030)
- Hypercarb Guard column (5 µm, 1 × 10 mm, Thermo Scientific, cat. no. 35005-011001) with Universal Uniguard Holder (1.0 mm i.d., Thermo Scientific, cat. no. 851-00)
- 96-Well twin.tec PCR plates (Eppendorf; Thermo Fisher Scientific, cat. no. E951020401)
- Polypropylene vial (250 µL; Agilent, cat. no. 5188-2788)
- Crimp/snap-top vials and caps (2 mL; Agilent Technologies, cat. no. 5182-0541)

- E-Pure water purification system (Thermo Fisher Scientific, cat. no. D4631)
- Ultra-high performance liquid chromatography (UHPLC) system (1290 Infinity II LC system, Agilent Technologies)
- Triple Quadrupole LC/MS (Agilent Technologies, model no. 6495A)
- High-performance liquid chromatography (HPLC) system (1260 Infinity II; Agilent Technologies)
- Accurate-mass Q-TOF LC/MS system (Agilent Technologies, model no. 6530)
- Incubator/Oven (Jeio Tech, model no. OF-01E)
- Centrifuge (Eppendorf, model no. 5811F)
- Metal clamp for 96-well plates (machined using pictures and dimensions in **Supplementary Figure 4.1**)

SOFTWARE

All required software can be run on a standard personal computer equipped with a Windows operating system.

- Agilent MassHunter Workstation for LC/QQQ (B.08.00; Agilent Technologies)
- Agilent MassHunter Workstation for LC/TOF and LC/Q-TOF (B.08.00; Agilent Technologies)
- Agilent MassHunter Workstation Qualitative Analysis (B.08.00 Agilent Technologies)
- Agilent MassHunter Workstation Quantitative Analysis (B.08.00; Agilent Technologies)

REAGENT SETUP

CRITICAL: Water generated from the E-pure water purification system (18 MΩ resistance) is used for reagent steps unless other types of water are specified

1-phenyl-3-methyl-5-pyrazolone (PMP) solution for monosaccharide and linkage analysis

This solution is a mix of equal parts methanolic 0.2 M PMP and ammonia solution (28-30 % w/v). To prepare enough for one 96-well plate of samples, weigh 522.6 mg into a 50 mL tube. Dissolve completely in 15 mL of methanol. Add 15 mL of ammonia solution and vortex well.

! CAUTION: Methanol is flammable and toxic. Ammonia is a skin and respiratory irritant. Prepare this solution in a fume hood.

Saturated NaOH solution for linkage analysis

Weigh 6.3 g of NaOH into a 15 mL Falcon tube. Add 5 mL of water and vortex until completely dissolved. This may be scaled down proportionately if less solution is needed.

! CAUTION: NaOH is corrosive, and a significant amount of heat is generated from preparing the saturated solution. Ensure the lid is securely tightened before vortexing. Periodically crack the lid to relieve excess vapor pressure.

- CRITICAL: This solution should be used immediately after preparation. Allowing it to cool will result in a solid or slurry.

Monosaccharide analysis LC solvent A

This solvent is 5% (vol/vol) LC-MS-grade ACN in water and 25 mM ammonium acetate with a pH of 8.2. To a 1-liter volumetric flask, add 50 mL ACN and add water to a final volume of 1 L. Weigh out approx. 1.927 g of ammonium acetate and dissolve with the prepared 1 L 5% ACN

mixture. Add approximately 150 µL ammonia solution to adjust the pH to 8.2. This solution can be used for up to one week.

- CRITICAL: This solution should be prepared fresh before running a batch of samples.

Monosaccharide analysis LC solvent B

This solvent is 95% (vol/vol) LC-MS-grade ACN in water. To a 1-liter volumetric flask, add 50 mL Milli-Q water and add ACN to fill in a final volume of 1 L. Transfer to an LC solvent container. This solution can be used for up to one week.

- CRITICAL: This solution should be prepared fresh right before running a batch of samples.

Glycosidic linkage analysis LC solvent A

This solvent is 5% (vol/vol) LC-MS-grade ACN in water and 25 mM ammonium acetate with a pH of 7.7. To a 1-liter volumetric flask, add 50 mL ACN and add water to a final volume of 1 L. Weigh out 1.927 g of ammonium acetate and dissolve with the prepared 1 L 5% ACN mixture. Add approximately 60 µL ammonium hydroxide to adjust the pH to 7.7. Transfer to an LC solvent container. This solution can be used for up to one week.

- CRITICAL: This solution should be prepared fresh right before running a batch of samples.

Glycosidic linkage analysis LC solvent B

This solvent is 95% (vol/vol) LC-MS-grade ACN in water. To a 1-liter volumetric flask, add 50 mL Milli-Q water and add ACN to fill in a final volume of 1 L. Transfer to an LC solvent container. This solution can be used for up to one week.

- CRITICAL: This solution should be prepared fresh right before running a batch of samples.

Polysaccharide (FITDOG) analysis LC solvent A

This solvent is 3% (v/v) of ACN in water with 0.1% (v/v) FA. In a 1 L volumetric flask, add 30 mL of ACN and 1 mL of FA, fill to mark with Milli-Q water, and mix thoroughly. Fill to mark again with Milli-Q water if needed. Transfer to an LC solvent container. This solution can be used for up to one week.

- CRITICAL: This solution should be prepared fresh right before running a batch of samples.

Polysaccharide (FITDOG) analysis LC solvent B

This solvent is 90% (v/v) of ACN in water with 0.1% (v/v) FA. In a 1 L volumetric flask, add 90 mL of Milli-Q and 1 mL of FA, fill to mark with ACN, and mix thoroughly. Fill to mark again with ACN if needed. Transfer to an amber glass LC solvent container. This solution can be used for up to one week.

- CRITICAL: This solution should be prepared fresh right before running a batch of samples.

Sodium acetate buffer for FITDOG reaction

This solution is 44 mM $NaCH_3COO$ in water at pH 5.2. Prepare 44 mM sodium acetate solution. Adjust pH to 5.2 by adding glacial acetic acid. This solution can be stored at 4 °C for several months.

FITDOG reaction mixture

This solution has 41.8 mM $NaCH_3COO$, 65 μM Fe(III), and 1.5% (w/v) H_2O_2. Mix 95 mL of sodium acetate buffer and 5 mL of 30% (w/v) H_2O_2. Weigh 3.56 mg of $Fe_2(SO_4)_3 \cdot 5H_2O$ and dissolve it in the prepared solution.

- CRITICAL: This solution should be prepared freshly each time, right before the experiment.

Quenching solution for FITDOG reaction

This solution is 2 M NaOH in water. Weight 4.0 g of NaOH and dissolve it in 50 mL water.

- CRITICAL: This solution should be prepared freshly each time, right before the experiment.

Reducing solution for FITDOG analysis

This solution is 1 M $NaBH_4$ solution. Weigh 1.89 g of $NaBH_4$ and dissolve it in 50 mL water.

- CAUTION: Dissolving $NaBH_4$ in water forms H_2 gas. Depressurize container a few times while dissolving $NaBH_4$.
- CRITICAL: This solution should be prepared freshly each time, right before the experiment.

PGC Priming Solution

This solution is 80% ACN with 0.1% TFA. To prepare 1 L of solution, mix 800 mL ACN, 1 mL TFA, and 199 mL of water. This solution can be stored at 4 °C for several months.

PGC Elution Solution

This solution is 40% ACN with 0.05% TFA. To prepare 1 L of solution, mix 400 mL ACN, 0.5 mL TFA, and 599.5 mL water. This solution can be stored at 4 °C for several months.

Pooled calibration standards for quantitative monosaccharide analysis

Weigh 14 mg of each of the 14 monosaccharide standards into separate 1.5 mL tubes. Add water to prepare solutions of exactly 14 mg/mL. Pool equal aliquots together to prepare a 1 mg/mL pooled stock solution. Serially dilute the 1 mg/mL pool according to the calibration levels in **Table 4.2**. Once prepared, the stock solution and calibration standards may be stored at –20 °C and kept for several months.

Table 4.2. Calibrators and concentrations (for each monosaccharide) used for monosaccharide analysis.

Level	Conc. (ng/mL)
L1	1
L2	10
L3	100
L4	1000
L5	5000
L6	10000
L7	25000
L8	50000
L9	75000
L10	100000

Pooled oligosaccharide standards for linkage analysis

Weigh 10 mg of each of the linkage standards into separate 1.5 mL tubes. Add water to prepare stock solutions of 10 mg/mL. Combine equal aliquots of each to create a pooled stock solution to be used directly for linkage analysis. Once prepared, the stock solutions may be stored at -20 ˚C and kept for several months.

Calibration standards for polysaccharide (FITDOG) analysis

Several polysaccharide standards may be pooled together to reduce the number of samples. A recommended pooling scheme is summarized in **Table 4.3**. Weigh polysaccharides into 2-mL screwcap tube and add 1 mL of water, incubate at 100 ^{o}C for 1 h, and then homogenize with stainless-steel beads. If needed, calibrator stock mixtures may be stored at -20 ^{o}C for several weeks. Serial dilution should be carried out immediately before the experiment.

Table 4.3. Scheme for pooling polysaccharide standards.

Pool 1	Pool 2	Pool 3	Pool 4
Arabinan	Xylan	Starch	Arabinoxylan
Galactan	Mannan	Cellulose	Galactomannan
Xyloglucan	β-glucan		
	Chitin		

PROCEDURE

Sample Preparation

Lyophilization (freeze-drying) and dry homogenization of samples *Timing: 16-72 h, depending on the moisture content of the samples.

1. Collect the sample, place it in an appropriately sized screw cap tube, and transfer to a -80 °C freezer for at least 3 h. If moisture content is needed, record the mass of the samples and the tubes.
2. Start the freeze-dryer by turning on the condenser. It will take several minutes for the temperature of the condenser tray to reach the appropriate temperature (–60 ^{o}C). Once the condenser tray is cold enough, start the vacuum.

3. Remove the frozen sample tubes from the freezer, remove the lids, and place them gently back on top of their respective tubes.
    - CRITICAL: Ensure the cap is completely free from the thread on the tube as the loss of pressure may cause the tube to seal.
4. Gently place the frozen sample tubes in the glass freeze-dryer jars, fasten them to the vacuum manifold, and lyophilize until a minimum moisture content is achieved. This procedure can be accomplished by weighing the sample tube at intervals. However, we have found that 3 d is sufficient to dry nearly all sample types without the necessity of weighing. If moisture content is needed, record the mass of the dried sample.
5. If the samples are powdered after drying, they may be homogenized directly with a bead mill (step 6). If samples are oily (like nut butters) or appear hygroscopic (like many fruits), they must be flash-frozen with liquid nitrogen and homogenized with a mortar and pestle.
6. Transfer the lyophilized samples into separate 30 mL screw-cap tubes and grind the samples using a bead mill homogenizer with 5 mm stainless-steel beads for 2 min at 4 m/s.
    - CRITICAL: Ensure samples are a homogenous powder before moving forward.

Preparation of sample suspensions *Timing: 2 h

7. If removal of low-molecular weight saccharides (mono-, di-, and oligosaccharides) before compositional analysis is desired, follow option A. Otherwise, follow option B.
    - **A. Removal of low-molecular weight saccharides by ethanol precipitation *Timing: 3 h**
        i. Weigh out 10 mg (± 0.5 mg) of each sample into 1.5 mL screw-cap tubes using an analytical balance and record the mass. Add 1 mL of 80 % EtOH.

ii. Homogenize the samples on a bead mill at 4 m/s for 1 min. Centrifuge at 10,000 g for 10 min. Carefully remove the supernatant without disturbing the pellet using a pipette, add another 1 mL of 80% EtOH, homogenize, and centrifuge again to wash the pellet. Repeat the wash once more.

iii. Remove the supernatant from the final wash and dry the resulting pellets completely in a centrifugal vacuum evaporator. This takes approximately 1 h, depending on how much of the supernatant was successfully removed.

- PAUSE POINT: The dried pellet may be stored at –20 °C until further preparation.

iv. Add 1 mL of water and 2-mm stainless-steel beads to the sample pellets.

v. Homogenize the samples on a bead mill at 4 m/s for 2 min. Incubate the suspended samples at 100 °C for 1 h before bead milling once more with the same settings.

- PAUSE POINT: The suspended sample stocks may be stored at –20 °C before proceeding. If frozen, samples should be homogenized again via bead mill after thawing.

- **B. Preparation of sample stock suspensions without EtOH precipitation**

***Timing: 2 h**

i. Weigh out 10 mg (± 0.5 mg) of the samples into 1.5 mL screw-cap tubes using an analytical balance and record the mass. Add 1 mL of water and 2 mm stainless-steel beads.

ii. Homogenize the samples on a bead mill at 4 m/s for 2 min. Incubate the suspended samples at 100 °C for 1 h before homogenizing once more with the same settings.

- PAUSE POINT: The suspended sample stocks may be stored at -20 °C before further analysis. If frozen, samples should be homogenized again via bead mill after thawing.

Quantitative Monosaccharide Analysis: Acid Hydrolysis * Timing: 1.5 h

8. Aliquot 10 µL from the sample stock suspension into each well of a 96-well plate or1.5 mL screw cap tubes. Add 90 µL water to each sample.
    - CRITICAL: Stock suspensions should be vortexed before pipetting to ensure homogeneity.
9. Add 44.5 µL of TFA to each well/tube. Seal the plate or tubes immediately to avoid evaporation. Vortex lightly using a plate shaker/vortex mixer for 1 min and centrifuge for 30 s at 300 × g.
    - CAUTION: TFA is a corrosive chemical. Use only in a fume hood.
10. After centrifuging, incubate the 96-well plate/tubes at 121 °C for 1 h.
    - CRITICAL: If using a 96-well plate, the plates must be sealed well using a clamp. These can be machined quite easily (see pictures and dimensions in **Figure 4.3**).
11. Once the incubation is complete, remove the plate from the oven, and allow it to cool to room temperature. Once cooled, remove from the clamp and centrifuge at 300 × g for 30s.
    - CAUTION: The samples and clamps are extremely hot after incubation. Handle only with heat-resistant gloves.
12. Add 855.5 µL of ice-cold water. Centrifuge at 300 × g for 30 s.

Quantitative Monosaccharide Analysis: PMP Derivatization * Timing: 45 min

13. Transfer 10 µL of the hydrolyzed sample solution to another 96-well plate or screw cap tube. Transfer 50 µL of each level (L1-L10) of the monosaccharide calibration curve.
14. Transfer 200 µL of the PMP solution to each well or tube.
15. Seal the plate or tubes, vortex lightly for 1 min, and centrifuge at 300 × g. Incubate at 70 °C for 30 min.
    - Critical Step: 96-well plates must be sealed and incubated with the clamp during the reaction.
16. Once the incubation is complete, remove the samples from the incubator, and allow to cool to room temperature. Centrifuge at 300 × g for 30 s.
    - CAUTION: The samples and clamps are hot after incubation. Handle only with heat-resistant gloves.
    - CRITICAL: 96-well plates should remain securely sealed while cooling and should be centrifuged before removing the seal.
17. Once cooled, remove the lids, place the samples in a centrifugal vacuum evaporator fitted with a pressure programmer and dry completely. This will take at least overnight to dry.
    - CRITICAL: The dryer should be programmed for the stepwise evaporation of the MeOH/ammonia/water solution to avoid solvent bumping. We have optimized this to include an initial pressure drop from atmosphere to 200 mbar followed by a gradient drop to 10 mbar over 4 h.
    - PAUSE POINT: The dried and derivatized samples may be stored at -20 °C for several weeks before subsequent steps.

Quantitative Monosaccharide Analysis: Chloroform Extraction *Timing: 30 min

18. Add 250 µL chloroform into each tube or well of the 96-well plate containing samples and calibration standards. Vortex on a plate shaker until the pellet is nearly completely dissolved. Add 250 µL water, vortex for 1 min on a plate shaker, and centrifuge at 300 × g for 1 min.
    - CAUTION: Chloroform is flammable, toxic, and will dissolve most disposable solvent reservoirs. Perform these steps in a fume hood with a glass solvent reservoir.
19. Remove and discard 150 µL of the chloroform layer (bottom) from each well. Add another 250 µL chloroform into each well. Vortex for 1 min and centrifuge at 300 × g.
    - CAUTION: Perform this extraction in a fume hood and discard waste in an appropriate hazardous waste vessel.
20. Transfer 100 µL of the top aqueous layer from each well into a 96-well analysis plate or autosampler vials compatible with the autosampler to be used. Centrifuge before injection.
    - CRITICAL: Ensure that only the aqueous layer is transferred.

Quantitative Monosaccharide Analysis: UHPLC-QqQ MS analysis *Timing: Depends on batch size (~5 min per sample)

21. Start the data acquisition software (e.g., Agilent MassHunter) on the UHPLC-QqQ MS and place the 96-well plate or vials into the autosampler compartment.
    - CRITICAL: At the beginning of the batch, run at least 3-5 blanks and a mid-level calibration standard to equilibrate the LC system. For instrument quality control (QC), a mid-level calibration standard should be injected every 12 samples along with a blank sample.

! Troubleshooting

22. Make a worklist using the "Worklist" Tab or load a worklist created in the Offline Worklist Editor.
23. Load the freshly prepared solvents A and B for monosaccharide analysis onto the pump. Update the solvent level in "Bottle Filling" section in the software. Purge both solvents for at least 5 min at a flowrate of 5 mL/min. A detailed list of the LC-MS parameters is included in **Table 4.4**.

Table 4.4. LC-MS/MS data-acquisition parameters for monosaccharide, linkage, and polysaccharide analysis.

	Polysaccaride (FITDOG)	**Monosaccharide**	**Linkage**
LC parameters			
Column packing material	PGC	C18	C18
Typical injection volume (uL)	10	2	2
Solvent A (vol/vol)	0.1% FA, 3% ACN, 96.9% H_2O	25 mM ammonium acetate in 5% ACN (pH 8.2)	25 mM ammonium acetate in 5% ACN (pH 7.7)
Solvent B (vol/vol)	0.1% FA, 90% ACN, 9.9% H_2O	95% ACN, 5% H_2O	95% ACN, 5% H_2O
Flow rate (mL/min)	0.132	1.05	~ 0.20
Gradient (% B)	0-15 min: 3%-25%	0-1.9 min: 11%	0-5 min: 21%
	15-18 min: 25%	1.9-2.2 min: 11%-99%	5-9 min: 21%-22%
	18-30 min: 25%-99%	2.2-3.8 min: 99%	9-11 min: 22%
	30-32 min: 99%	3.8-4.6 min: 11%	11-13.6 min: 22-24.5%
	32-34 min: 99%-3%		13.6-13.8 min: 99%
	34-45 min: 3%		13.8-16 min: 21%
ESI source parameters			
Polarity	Positive	Positive	Positive
Drying gas temperature (^oC)	100	290	290
Drying gas flow (L/min)	9	11	11

Sheath gas temperature (^{o}C)	150	300	300
Sheath gas flow (L/min)	11	12	12
Nebulizer (psi)	20	30	30
Capillary voltage (V)	1800	1800	1800
Nozzle voltage (V)	1500	1500	1500
Fragmentor (V)	65		
Skimmer (V)	50		
Oct 1 RF Vpp (V)	500		
High pressure RF (V)		150	150
Low pressure RF (V)		60	60
MS parameters		See MRM table in Supplementary Table 3	See MRM table in Supplementary Table 4
m/z range	(MS) 250-3000, (MS/MS) 50-2000		
Cycle time	4.86 s		
Acquisition mode	Auto MS/MS (DDA)		
MS scan rate	1 spectra/s (1000 ms/spectrum)		
MS/MS scan rate	1.33 spectra/s (752 ms/spectrum)		
MS threshold	Absolute threshold, 50; relative threshold, 0.01%		
MS/MS threshold	Absolute threshold, 5; relative threshold, 0.01%		
Calibrant ion	922.009798 (± 35 ppm)		
Activation type	CID		
Activation energy	CE = 1.45 × (*m/z*) / 100 - 3.5		
Max. precursors per cycle	5		
Precursor selection threshold	Absolute threshold, 50; relative threshold, 0.01%		

Precursor target (counts/spectrum)	25,000		
Dynamic exclusion	Excluded after 2 spectra, released after 0.5 min		
Precursor charge state preferemce	2, 1, unknown, 3, >3		
Isotope model	Common organic molecules		

24. Install an Agilent Poroshell HPH C18 column (2.1 mm × 50 mm, 1.8 μm particle size) equipped with an Agilent Poroshell HPH C18 guard cartridge (2.1 mm × 5 mm, 1.8 μm particle size) into the column compartment. After purging, turn on the pump and allow to equilibrate for at least for 10 min.
    - CRITICAL: Remember to update the column position after installing it. During the conditioning, start from a lower flowrate like 0.2 mL/min and increase stepwise to match the operating flowrate.
25. Start the worklist after purging, conditioning, and monitoring the first QC using MassHunter Qualitative B.08.00 software. Extract the transactions for the 14 monosaccharides and assign their retention time using **Table 4.5**. Adjust the retention time of the dMRM transitions in the Acquisition tab, if needed.

! Troubleshooting

- CRITICAL: Observe the QCs throughout the run to see if they are reproducible in terms of retention time and abundance.
- PAUSE POINT: After finishing the batch, the plate or tubes can be stored at –20 °C. Samples can be stored for up to 2 weeks for re-injection. Wrap well with aluminum foil before doing so.

Table 4.5. Dynamic MRM transition list for the targeted monosaccharide analysis using UPLC-QqQ.

Name	TS	Transition	Scan	Type	RT	Left RT Delta	Right RT Delta	RT Delta Unit	Ion Polarity
Fructose	1	511.2 -> 175.1	MRM	Target	0.269	0.1	0.1	Minutes	Positive
Mannose	1	511.2 -> 175.1	MRM	Target	0.753	0.1	0.1	Minutes	Positive
Allose	1	511.2 -> 175.1	MRM	Target	0.899	0.1	0.1	Minutes	Positive
Ribose	1	481.2 -> 175.1	MRM	Target	0.901	0.2	0.2	Minutes	Positive
GlcA	1	525.2 -> 175.1	MRM	Target	0.915	0.1	0.1	Minutes	Positive
Rhamnose	1	495.2 -> 175.1	MRM	Target	0.986	0.2	0.2	Minutes	Positive
GalA	1	525.2 -> 175.1	MRM	Target	1.051	0.1	0.1	Minutes	Positive
Glucose	1	511.2 -> 175.1	MRM	Target	1.491	0.2	0.2	Minutes	Positive
GlcNAc	1	552.2 -> 175.1	MRM	Target	1.503	0.2	0.2	Minutes	Positive
GalNAc	1	552.2 -> 175.1	MRM	Target	1.657	0.2	0.2	Minutes	Positive
Galactose	1	511.2 -> 175.1	MRM	Target	1.695	0.2	0.2	Minutes	Positive
Xylose	1	481.2 -> 175.1	MRM	Target	1.723	0.2	0.2	Minutes	Positive
Arabinose	1	481.2 -> 175.1	MRM	Target	1.816	0.2	0.2	Minutes	Positive
Fucose	1	495.2 -> 175.1	MRM	Target	2.035	0.2	0.2	Minutes	Positive

Quantitative Monosaccharide Analysis: Data Analysis * Timing: 30 min – 60 min

26. Start the MassHunter Quantitative Analysis B.08.00 software and open a new batch under the data file folder. In the window "Add Samples," select monosaccharide calibration curve standards and samples and click "OK".
27. After loading samples, label calibration standards with their Level Name as listed in **Table 4.2** and update their type as "Cal" for calibration.
    - CRITICAL: If the columns for "Level" and "Type" are not shown, right-click to "Sample" column to add. Sample type can be modified in the worklist as well.
28. Click 'Method" --> "New"--> "New Method from Acquired MRM Data" and select one of the mid-level calibration standards from this batch. Click the "MRM Compound Setup" tab and a list of compound names, transitions and RT will be shown. Update the list of 14 monosaccharide compounds.
29. Assign 14 monosaccharides based on the elution order of isomers shown in **Table 4.5**. For isomers that are monitored using the same dMRM transitions, select the compound and right-click to "Duplicate compound," update the compound name, and assign the retention time.
30. Click "Qualifier Setup" tab and make sure the quantifier and qualifier transitions match with **Table 4.5**. For each compound, the quantifier has a product ion of 175.2 *m/z* and a qualifier ion of 217.1 *m/z* (GlcNAc/GalNAc are the exception with qualifier ions of 258.1 *m/z*. The precursor ion should match for both. If any compound does not have a qualifier, right-click the compound and add "New Qualifier" manually.
31. Click "Concentration Setup," select a compound to add "New Calibration Level". Add 10 calibration levels and update the "Level" and "Conc." sections as shown in **Table 4.2**.

Apply the calibration level to all compounds. After setting the calibration curve for one compound, right-click and select "Copy Calibration Level to…" and select all compounds.

32. Save as a new method and click exit to "Analyze" the batch. This method can be saved for future analysis requiring that the user need only update the retention times of the compounds.
33. Go to "View" and click "Compounds-at-a-glance." A window displaying the integrated peaks for each compound in each sample will come up. Check the integrations of each compound and correct if needed.
34. The calibration curve for each monosaccharide is displayed in the "Calibration Curve" box. Exclude high-end data points, if necessary, based on the r^2 value. Retain at least 6 points. Click "Analyze Batch" again to apply any corrected integration in the calibration curves to the samples.

! Troubleshooting

35. Go to "File" and export the table as an Excel (.xlsx) file.
    - CRITICAL: Ensure the "Final Conc." for each compound is exported. To do this, ensure "Multiple Sample/Compound View" is displayed in the software. Click "Add/Remove Columns" and only include the "Final Conc." column for export.
36. After exporting the table, convert the sample concentration to mg of monosaccharide per milligram of dried material (mg/mg). The default sample concentration unit expressed in the software is ng/mL. First, convert the sample units to mg/mL by dividing by 10^6. During the acid hydrolysis, each sample is 100-fold diluted. During the PMP derivatization, the sample is 5-fold diluted relative to the standard calibration curve. Therefore, a dilution factor of 500 is needed to calculate the sample concentration. Lastly, the sample

concentration is divided by the mass weighed to create its stock solution to arrive at a unit of mg/mg.

Glycosidic Linkage Analysis: Permethylation * Timing: 4 h

37. Pipette 5 µL from the 10 mg/mL sample stock suspension (approx. 50 µg sample) and the pooled linkage standards to a 96-well plate or 1.5 mL screw cap tube. Dry completely under vacuum centrifugation.

    - CRITICAL: Stock suspensions should be vortexed before aliquoting to ensure homogeneity.

38. Add 5 µL of the saturated NaOH solution to the 96-well plate or 1.5 mL tubes containing the samples. Replace the plate/tube lid and centrifuge at 300 g for 1 min. Place the 96-well plate/tubes on a plate shaker for 30 min.

    - CRITICAL: Ensure the 5 µL of NaOH solution is centrifuged down to the bottom of the well/tube to redissolve the sample.

39. Stop the shaker after 30 min and purge the 96-well plate/tubes with argon in the vacuum chamber.
40. Purge the DMSO bottle with argon and transfer to a glass solvent reservoir using a needle and syringe. Add 150 µL of argon-purged DMSO into each well. Centrifuge at 300 g and vortex for 30 min in the argon-purged chamber.

- CRITICAL: DMSO can be purged using a Schlenk line setup. Remove air in the bottle by vacuum and purge with argon. Use a polypropylene syringe and a clean cannula to transfer purged DMSO in a glass reservoir.
- CAUTION: DMSO is toxic and flammable. Use only in the fume hood.

! Troubleshooting

41. Add 40 µL iodomethane into each well/tube. Centrifuge at 300 × g for 30 s and vortex for 50 min on a plate shaker.

    - CAUTION: Iodomethane is toxic and an oxidizer. Use only in a fume hood.

42. Quench the reaction with 700 µL ice-cold water and 300 µL dichloromethane. Vortex and centrifuge at 300 × g for 1 min.
43. Remove and discard 700 µL of the water layer (top) and add another 700 µL fresh water to each well. Vortex and centrifuge at 300 × g for 1 min Repeat this step twice.

    - CRITICAL: Graduated 1-mL pipette tips help to visualize alignment during the water removal step when using a plate and multi-channel pipette.

44. After 3 extractions, remove as much of the water layer as possible without disturbing the DCM layer and dry the plate using vacuum centrifugation.

Glycosidic Linkage Analysis: Acid Hydrolysis *Timing: 2.5 h

45. Reconstitute the dried, permethylated sample with 60.5 µL water and vortex the samples for several minutes.
46. Add 30.5 µL of TFA and seal the plate/tubes immediately to avoid evaporation. Vortex for 1 min and centrifuge at 300 × g.

- CAUTION: TFA is a corrosive chemical. Use only in a fume hood.
- CRITICAL: If using a 96-well plate, the plates must be sealed well using a clamp. These can be machined quite easily (see pictures and dimensions in **Figure 4.3**).

47. Incubate the 96 well plate/tubes at 100 °C for 2 h.

    - CRITICAL: If using 96-well format, the plate must be tightly sealed with the clamp to avoid evaporation.

48. Once the incubation is complete, remove from the oven, and allow to cool to room temperature. Once cooled, centrifuge at 300 × g for 30 s and place the samples in a centrifugal vacuum evaporator. It usually takes 2-4 h to dry completely.

    - CAUTION: The samples and clamps are extremely hot at after incubation. Handle only with heat-resistant gloves.
    - CRITICAL: The lids (if using 96-well plate) should be securely sealed during the cool-down process and the plates should be centrifuged prior to opening.

Glycosidic Linkage Analysis: PMP Derivatization * Timing: 45 min

49. Transfer 200 µL of PMP solution to each well for derivatization.

    - CAUTION: MeOH is flammable and NH_4OH is corrosive. Work with this solution in a fume hood.

50. Seal the plate or tubes and vortex for 1 min and centrifuge at 300 × g. Incubate at 70 °C for 30 min.

    - CRITICAL: 96-well plates must be sealed and incubated with the clamp during the reaction.

51. Once the incubation is complete, remove the samples from the incubator, and allow them to cool to room temperature. Centrifuge at 300 × g for 30 s.

    - CAUTION: The samples and clamps are hot after incubation. Handle only with heat-resistant gloves.

52. Once cooled, place the samples in a centrifugal vacuum evaporator fitted with a pressure programmer (more details can be found in Step 17) and dry completely. This will take at least overnight.

    - PAUSE POINT: The dried and derivatized samples can be stored at -20 °C for several weeks until instrument analysis.

Glycosidic Linkage Analysis: UHPLC-QqQ-MS analysis * Timing: 16 min per sample

53. Add 70 µL MeOH and 30 µL Nanopure water into each well/tube. Vortex and centrifuge at 300 × g for 1 min at RT.

    - CRITICAL: Ensure methanol dissolves the dried pellet before adding Nanopure water.

54. Transfer 70 µL from each sample into an injection 96 well plate/injection vial. Centrifuge before injection.

55. Start the Acquisition software and place 96-well injection plate into the autosampler compartment and close the door probably.

    - CRITICAL: At the beginning of the batch, run at least 3-5 blanks and a calibration standard point to equilibrate the LC system. For QC, an

oligosaccharide/polysaccharide standard could be injected every 12th sample along with a blank sample.

! Troubleshooting

56. Make a worklist using the "Worklist" Tab or load the worklist created by the "Offline Worklist Editor"
57. Load the freshly prepared solvents A and B for monosaccharide analysis into the pump. Update the solvent level in "Bottle Filling" section in the software. Purge both solvents for at least 5 mins at 5 mL/min. The detailed glycosidic linkage analysis LC-MS method is included in **Table 4.4**.
58. Load the column with guard column in the LC compartment. After purging, turn on the pump and condition the column for 10 min.
    - CRITICAL: If using multi-column valve, update the column position after loading the column. If using a new column, condition the column using low flow rate. Start from a lower flowrate like 0.2 mL/min and increase stepwise to match the desired pressure (approx. 450 bar).
59. Start the worklist after purge and condition and monitor the first QC using quantitation software, such as MassHunter Qualitative B.08.00.

! Troubleshooting

60. Observe the QCs throughout the run to see if they are reproducible.
61. After finishing the batch, change the sealing mat and save the injection plate with aluminum wrap in –20 ^{o}C. The samples can be stored for 1 week for re-injection.

Glycosidic Linkage Analysis: Data analysis for glycosidic linkage * Timing: 1h, depending on the size of the batch

62. Start the mass spectrometry analysis software (e.g. MassHunter Qualitative Analysis B.08.00) and open the data files. Assign the retention time based on **Figure 4.4B** and **Table 4.1** by extracting each transition.

    - CRITICAL: Make sure QCs are reproducible and retention time shifts are minimal. Each glycosidic linkage has two or three transitions: if the product ion 217.2 is more abundant than 231.2, this linkage is most likely 2-linked; if the 231.2 is more abundant than 217.2, this linkage is most likely not 2-linked.

63. Start MassHunter Quantitative Analysis B.08.00 software and open a new batch under the data file folder.
64. When a window "Add Samples" pop out, select samples and click "OK".
65. Click "Method" --> "New" --> "New Method from Acquired MRM Data" and select one of the data files from this batch.
66. Click "MRM Compound Setup" tab and a list of compound names, transition and RT will be shown. Update the compound list from the previous assignment.

    - CRITICAL: Assign the retention time for each glycosidic linkage based on the elution order of isomers shown in **Table 4.1**.

67. Click "Qualifier Setup" tab, make sure the quantifier and qualifier match with the **Table 4.1**.

    - CRITICAL: For 2-linked glycosidic linkages, make sure the qualifier product ion is 217.1 *m/z*. Otherwise, the product ion for qualifier should be 231.2 *m/z*. The

precursor ion should match with each other. If the compound does not have a qualifier, right-click the compound and add “New Qualifier” manually.

68. Save as a new method and click exit to “Analyze” the batch. This method could be saved for future analysis and only update the retention time.
69. Go to “View” and click “Compounds-at-a-glance”. After a window popping out, check the integrations of each compound. Go to “File” and export Table as excel file.
    - CRITICAL: The “Area” for each compound should be exported. Enable the “Multiple Sample/Compound View” display in the software. Click “Add/Remove Columns” and only include “Area” column for export.
70. After exporting the table, rearrange the table order as needed. Relative composition of glycosidic linkage by peak area of each sample could be graphed using Excel.

Polysaccharide (FITDOG) Analysis: Depolymerization reaction * Timing: 2 h

71. Pipette 100 µL (approx. 1.0 mg sample) from the 10 mg/mL stock suspension of the sample to a 96-well plate or 1.5 mL screw cap tube. For calibrator standards, the same volume, 100 µL, is used for analysis.
    - CRITICAL: Stock suspensions should be vortexed before aliquoting to ensure homogeneity.
72. Add 900 µL of freshly made FITDOG Reaction Mixture to each well/tube. Mix the reaction mixture using pipette (for plate), or vortex (if using tubes).
73. Seal the plate or tubes and incubate at 100 ^{o}C for 45 min.

- CRITICAL: If using 96-well format, the plate must be tightly sealed with the clamp to avoid evaporation.

74. Remove samples from the oven and allow to cool to room temperature for around 10 min.
    - CAUTION: The samples and clamps are extremely hot at after incubation. Handle only with heat-resistant gloves.
75. Slowly add 500 µL of the quenching solution (2 M NaOH, freshly made). Slowly mix with repeated pipetting.
    - CRITICAL: Some bubbles may form so the quenching solution should be added as slow as possible to prevent cross well contamination.
76. Slowly add 61 µL of glacial acetic acid. Slowly mix with repeated pipetting.
    - CRITICAL: Some bubbles may form so the acetic acid should be added as slow as possible to prevent cross well contamination.
    - CRITICAL: Glacial acetic acid is corrosive. Use only in a fume hood.
77. Transfer 400 µL into a new clean 96-well plate, or tubes, for the reduction.

Polysaccharide (FITDOG) Analysis: Oligosaccharide reduction * Timing: 2 h

78. Slowly add 400 µL of the reducing solution (1 M $NaBH_4$, freshly made) into each well/tube.
    - CRITICAL: Bubbles will form so the reducing solution should be added as slow as possible to prevent cross well contamination. Cooling down the samples and the reducing solution with ice bath prior to mixing will minimize the formation of bubbles.
79. Loosely place the plate sealing mat or tube cap. Do not seal the lids as gas is formed during the reaction.

80. Incubate in the oven at 65 ^{o}C for 1 h.
81. After oven incubation, remove from oven and cool down to room temperature.
    - CAUTION: The samples will be hot at after incubation. Handle only with heat-resistant gloves.

Polysaccharide (FITDOG) Analysis: Solid phase extraction (SPE) * Timing: 2 h

All the subsequent centrifugation steps should be at 1000 × *g* for 1 min, unless otherwise stated.

82. Prime the C18 SPE plate by adding 250 µL ACN to each well and then centrifuge. Repeat once. Discard the washings.
83. Condition the C18 SPE with 250 µL water and then centrifuge. Repeat this step three more times. Discard the washings.
84. Transfer the C18 SPE plate to a clean collection plate. Load 400 µL of the reduced oligosaccharide sample. Centrifuge and collect the flow-through.
    - CRITICAL: Make sure that the SPE sits on a clean collection plate before loading the sample.
85. Load the remainder of the reduced sample, centrifuge, and collect flow-through. This is the end of the C18 SPE.
86. Prime the PGC SPE plate with 400 µL of each solution in the following order: (1) water, (2) PGC priming solution (80% ACN/ 0.1% TFA), (3) water. Centrifuge between each step and discard washings.
87. Load 400 µL of the C18-cleaned sample. Centrifuge and discard flow-through. Load the remainder of the sample.

88. Wash the bound oligosaccharides in the PGC plate with 400 µL water. Centrifuge and discard washing. Repeat this step 5 more times.

! Troubleshooting

89. Change the collection plate to a clean one. Elute the oligosaccharides with 400 µL of the PGC elution solution (40% ACN/ 0.05% TFA) and centrifuge at 1000 × *g* for 2 min.
    - CRITICAL: Make sure that the SPE sits on a clean collection plate before adding the elution solution.
90. Place the samples in a centrifugal vacuum evaporator to dry completely. The drying will take at least 12 h.
    - PAUSE POINT: The dried samples may be stored at –20 ^{o}C for several weeks before the subsequent steps.

Polysaccharide (FITDOG) Analysis: HPLC-qToF MS analysis *Timing: 45 min per sample

91. Reconstitute samples with 50 µL water and vortex mix for 15 min. Centrifuge at 1000 × *g* for 1 min.
92. Transfer 50 µL into LC plate or vials. Centrifuge at 1000 × *g* for 1 min.
93. Start the MassHunter Acquisition software on the HPLC-qToF MS and place the sample plate or vials in the autosampler compartment.
94. Make a worklist sequence using the “Worklist” tab in the acquisition software or in the Offline Worklist Editor software.
    - CRITICAL: It is recommended to run a sequence of blanks (at least 2) first, then a standard to check the performance of the instrument. A mixture of starch and cellulose can be used as a workflow QC. Calibrator standards are also

recommended to be injected first before the samples. Inject instrument QC every 10-12 samples.

95. Load the freshly made solvent A and B for polysaccharide (FITDOG) analysis. Make sure to update the solvent level in the "Bottle Filling" section in the software. Purge pump for at least 10 min at a flow rate of 5 mL/min at 50% A/ 50% B composition.
96. Install a Thermo Scientific Hypercarb PGC column (1 mm × 150 mm, 5 µm particle size) equipped with guard column (Hypercarb, 1 mm × 10 mm, 5 µm particle size) in the column compartment.
    - CRITICAL: New columns should be conditioned first, starting at a lower flow rate and increasing stepwise to match the operating flowrate (0.132 mL/min).
97. Start the worklist sequence after purging and conditioning the LC column. The complete parameters for the method are summarized in **Table 4.4**. Monitor the backpressure during the first run and check the oligosaccharide peaks from an initial QC run using MassHunter Qualitative software.

! Troubleshooting

98. Throughout the run, monitor QC signals in terms of retention times and ion count abundances.

! Troubleshooting

Polysaccharide (FITDOG) Analysis: Data analysis *Timing: depends on the batch size

99. Open MassHunter Qualitative Analysis software and load the LC-MS/MS files.

100. Using the calibrator standards, extract ion chromatograms (XIC) of each relevant oligosaccharide precursor *m/z* values and take note of the retention times. Refer to **Table 3.1** (Chapter 3) for the complete oligosaccharide library.
    - CAUTION: Confirm monosaccharide class compositions from the tandem mass spectra.
101. Open MassHunter Q-TOF Quantitative Analysis software and open a new batch in the same directory as the files. In the window “Add Samples”, include all relevant sample files and calibrator standards.
102. Edit the method (“Method” --> “Edit”, or F10) and update the compound and retention times.
103. Save as a new method and click “Exit” and “Analyze” the batch.
    - CRITICAL: This method can be saved for future analysis requiring that the user need only update the retention times of the compounds.
104. In the main window, click “View” and then “Compounds-at-a-glance" to view all XICs. Check the integrations of each compound and sample and correct them as needed.
105. After doing the peak integrations, close the “Compounds-at-a-glance" window, save the batch file, and then export the peak area table as .xlsx or .csv file.
106. Using Excel or other spreadsheet software, quantitation can be done on the peak areas. For each polysaccharide, identify the top 3 most abundant oligosaccharides, get the average peak area of these 3 oligosaccharides, and plot it against the initial concentration (µg/mL or mg/mL) used in the workflow.
    - CRITICAL: At least 5 points should be used for the calibration curve. Linear or quadratic fit can be used to generate the calibration curve for each polysaccharide.

107. Apply the top 3 averaging method to the samples and use the calibration curves to interpolate the quantities of polysaccharides present in the sample. Concentrations can be converted into mg/mg units by dividing by the weighed mass of the sample in mg.
108. Convert mg/mg dry basis to fresh weight basis by using the moisture content of the sample.

Timing

- Step 1-6, lyophilization and dry homogenization of samples: 16-72 h, depending on the moisture content of samples
- Step 7A, removal of low-molecular weight saccharides by ethanol precipitation and preparation of sample stock suspensions: 3 h, depending on the size of the batch
- Step 7B, preparation of sample stock suspensions without precipitation: 2 h, depending on the size of the batch
- Step 8-12, monosaccharide acid hydrolysis: 1.5 h
- Step 13-17, monosaccharide PMP derivatization: 45 min
- Step 18-20, monosaccharide chloroform extraction: 30 min
- Step 21-25, monosaccharide UHPLC-QqQ MS analysis, 5 min per sample
- Step 26-36, monosaccharide data analysis, 30-60 min
- Step 37-44, linkage permethylation: 4 h
- Step 45-48, linkage acid hydrolysis: 2.5 h
- Step 49-52, linkage PMP derivatization: 45 min
- Step 53-61, linkage UHPLC-QqQ MS analysis: 16 min per sample
- Step 62-70, linkage data analysis: 1 h, depending on the size of the batch
- Step 71-77, FITDOG depolymerization reaction: 2 h

- Step 78-81, FITDOG oligosaccharide reduction: 2 h
- Step 82-90, FITDOG SPE clean-up: 2 h
- Step 91-98, FITDOG HPLC-qTOF analysis: 45 min per sample
- Step 99-108 FITDOG data analysis: 12-16 h per 96-well plate

Troubleshooting

Troubleshooting advice is summarized in **Table 4.6**.

Table 4.6. Troubleshooting Table

Step	Problem	Possible Reason	Solution
25	No signal observed or peaks appear cut-off in monosaccharide analysis chromatograms	Retention-time shifting has caused compounds to fall outside the dMRM windows or autosampler is not set-up properly.	Adjust retention-time windows to reflect updated retention times in the instrument method. Ensure the sample plates are oriented correctly and the autosampler is set to draw from the bottom of the sample well
34	Calibration curves in monosaccharide analysis are not linear	Standards/samples were not derivatized properly.	Re-aliquot hydrolyzed samples along with standards, derivatize, and run again
40	Permethylation is not working or signficant underpermethylation is observed (high trisecting linkage abundances)	Samples were not adequately dissolved or the permethylation reaction was not carried out properly.	Prepare and perform linkage analysis on a fresh set of samples
55	No signal observed in the linkage analysis chromatograms	Sample plates are not oriented properly in the autosampler correctly or the autosampler is not set-up properly.	Ensure the sample plates are oriented correctly and the autosampler is set to draw from the bottom of the sample well
60	Retention times have signifcantly changed since last run or peaks are coeluting	The flowrate in the method may need to be changed or the LC solvent is old or not at the correct pH.	Adjust the flowrate in the method if needed. Ensure the LC solvent is freshly prepared, and the pH is around 7.7
88	Some of the sample or water is left in the SPE wells after centrifugation.	This sometimes happen due to the hydrophobic nature of the PGC.	Centrifuge for the second time at higher speed (1300 × g).
97	LC pump pressure is abnormally high or is fluctuating.	This can be caused by a clog in the LC tubings or in the column.	Identify the source of the clog starting from the tubes from the pump to the column. Guard column may also be replaced.
98	Retention times are shifting.	Possibly caused by clogged, fouled, or old guard or main column.	Flush the column with multiple alternating rounds of high aqueous and high organic compositions. Guard

			and/or the main column may be replaced.
98	Ion signals are fluctuating or decreasing throughout the batch run.	This may indicate problems with the ESI source or the QTOF instrument.	ESI may be visually checked and cleaned to minimize background noise signal. Check tune can be done to quickly assess the performance of the instrument and to re-calibrate the m/z axis.

ANTICIPATED RESULTS

Dietary fiber consumption of human microbiota-colonized mice

Extensive research on the human gut microbiome has been enabled by using microbiota-colonized gnotobiotic mouse models. In this approach, germ-free mice are colonized by a consortium of bacteria (from cultures, or fecal donors). By applying this protocol to these kinds of studies, we were able to monitor the consumption of the fibers by the gut microbes.[13, 14, 45, 46] Analysis of fecal samples of gnotobiotic mice showed that the linkages reflected the compositions of the fibers fed to the mice (**Figure 4.5A**). However, a comparison of microbe-colonized and germ-free mice showed specific linkages were drastically decreased in the presence of gut microbes (**Figure 4.5B**). Specifically, total arabinose (**Figure 4.5C**) and the t-*f*-arabinose linkage (**Figure 4.5D**) were decreased with microbe-colonized mice in each of the fiber supplement types. The glycomic data were further integrated with metagenomic and metatranscriptomic analysis to provide specificity for the glycosyl hydrolases thereby elucidating specific fiber-microbe-host interactions.

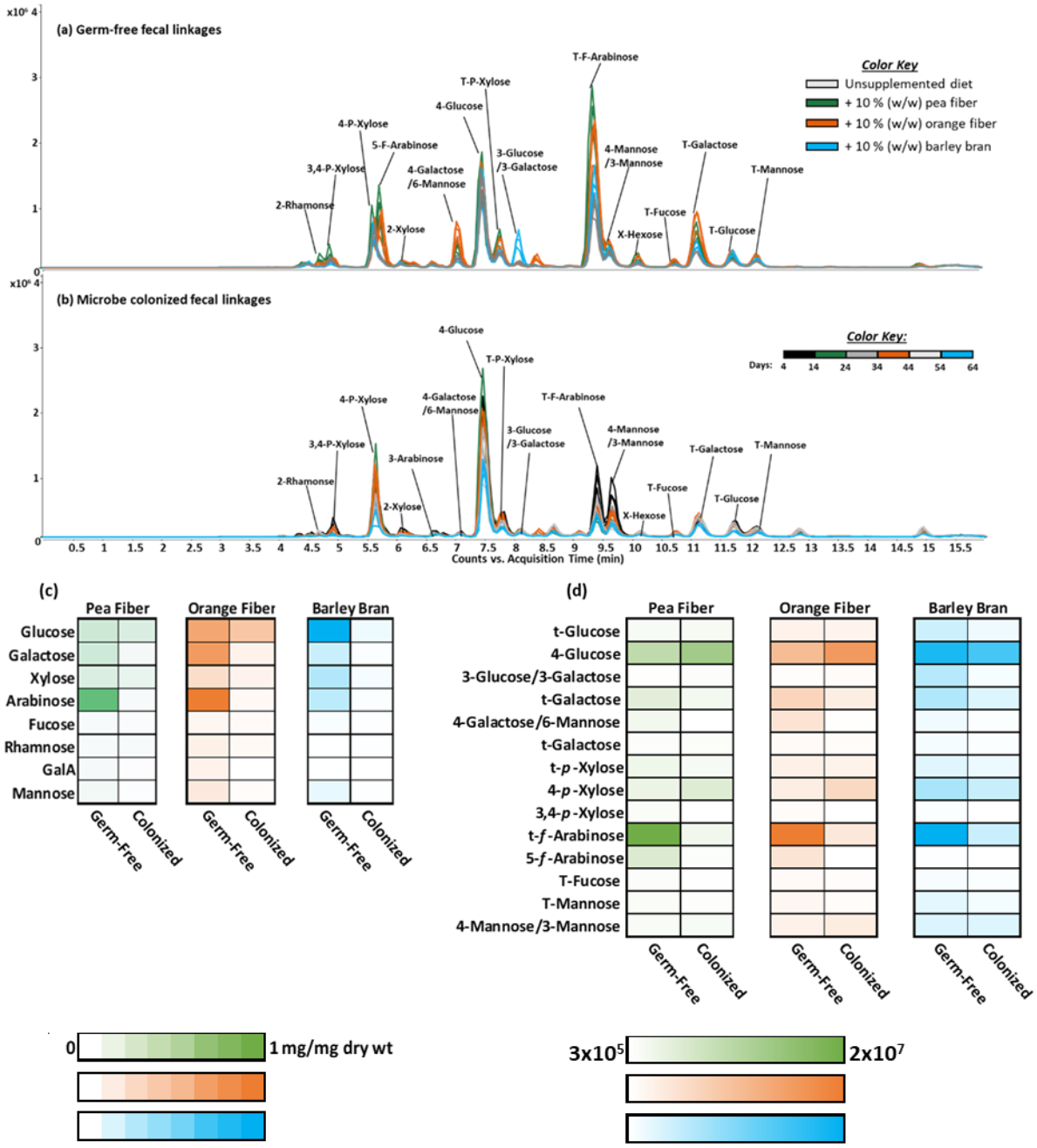


Figure 4.5. Monosaccharide and linkage profiles of fecal glycans obtained from germ-free and human-microbiota-colonized gnotobiotic mice. (a) Chromatogram of fecal glycosidic linkages from germ-free mice fed diets supplemented with 10 % (w/w) pea fiber, orange fiber, or barley bran. (b) Chromatogram of fecal linkages from inoculated mice fed the same diets over the course of 64 days. Replicate traces indicate technical replicates. (c) Heat maps illustrating the differences in the concentration of the most abundant fecal monosaccharide residues between germ-free and colonized mice expressed in mg/mg dry weight. (d) Differences in the observed peak area abundances of selected linkages between germ-free and colonized mice.

Differences in the glycomics profiles of apple varieties

Most crops have been bred or engineered to optimize yield, nutrient profile, or organoleptic properties, contributing to differences in glycan composition between varieties/ cultivars. The multi-glycomics workflow was used to characterize the carbohydrates present in different varieties of apples (**Figure 4.6**). Five different varieties were included, and for each variety, eight retail samples were analyzed. From the monosaccharide analysis, Granny Smith had a significantly higher amount of arabinose and galactose compared to the other four varieties (**Figure 4.6A, D**). The results were corroborated by the polysaccharide analysis where galactan was found higher in abundance in Granny Smith apples (**Figure 4.6C**). Arabinan was also found to be high in Granny Smith but the differences did not reach statistical significance (**Figure 4.6F**). Additionally, linkage analysis confirmed both findings from the monosaccharide and polysaccharide results. Namely, 4-galactose which is present primarily in galactans was significantly higher in Granny Smith (**Figure 4.6B**).

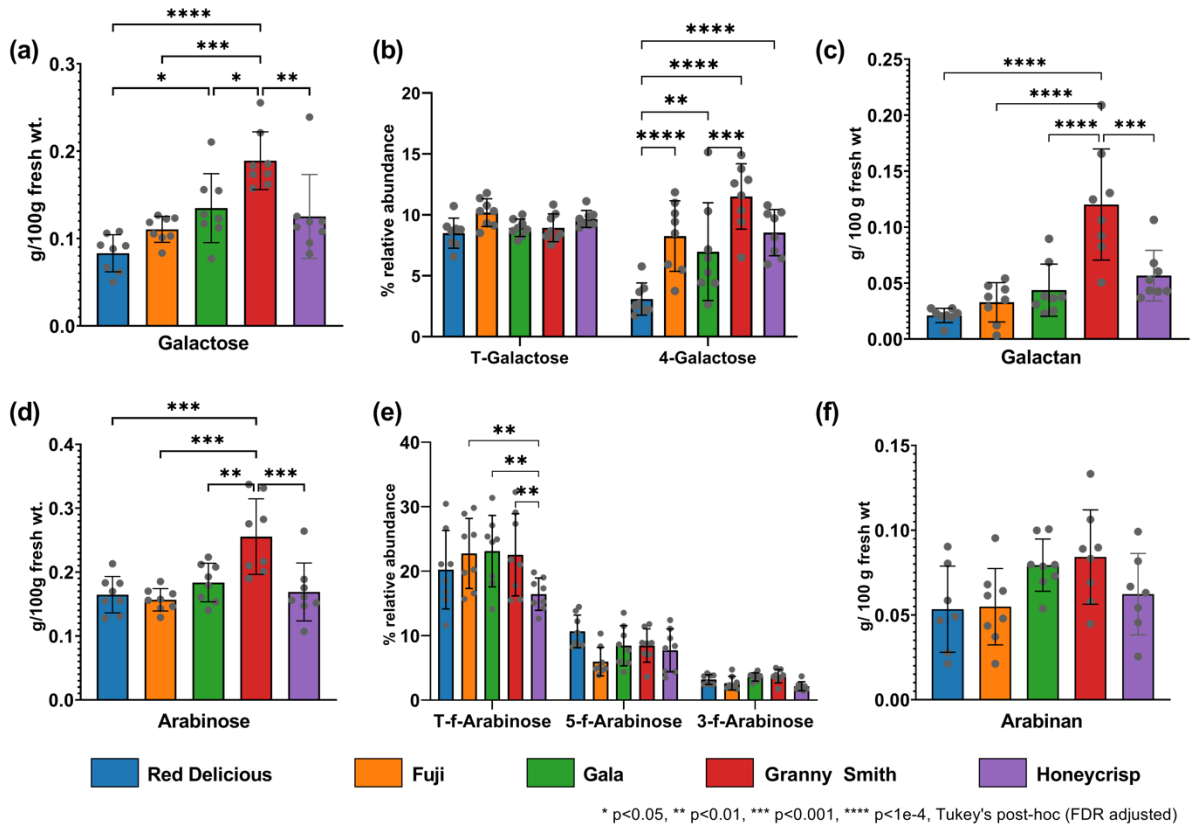


Figure 4.6. Monosaccharide, linkage and polysaccharide (FITDOG) composition analyses of different varieties of apples. Each bar represents mean; error bars represent standard deviation (n = 8 retail samples per variety). (a) Absolute galactose content measured using monosaccharide analysis. (b) Relative abundance of galactose linkages. (c) Absolute galactan content measured using FITDOG analysis. (d) Absolute arabinose content measured using monosaccharide analysis. (e) Relative abundance of arabinose linkages. (f) Absolute arabinan content measured using FITDOG analysis.

Carbohydrate-centric food database (Glycopedia)

One of the major attributes of this workflow is its increased throughput. This protocol is amenable to a 96-well plate format, enabling the parallel analysis of large batches of samples. Recently, we have published a carbohydrate-focused food composition database (Davis Food Glycopedia).[38] Over 800 food samples were analyzed for their monosaccharide compositions. Foods were categorized and clustered based on their monosaccharide profiles. For example,

grain products had significantly higher amounts of glucose. Conversely, fruits and vegetables had greater monosaccharide diversity (**Figure 4.7**). A powerful use for this database is in formulating diets and menus that can be tailored toward specific monosaccharide compositions. For example, consuming more whole grain products (vs. highly refined and processed grain products) will result in higher consumption of arabinose and xylose. We are continuously expanding this database in terms of both the number of food entries, as well as the depth of analysis.

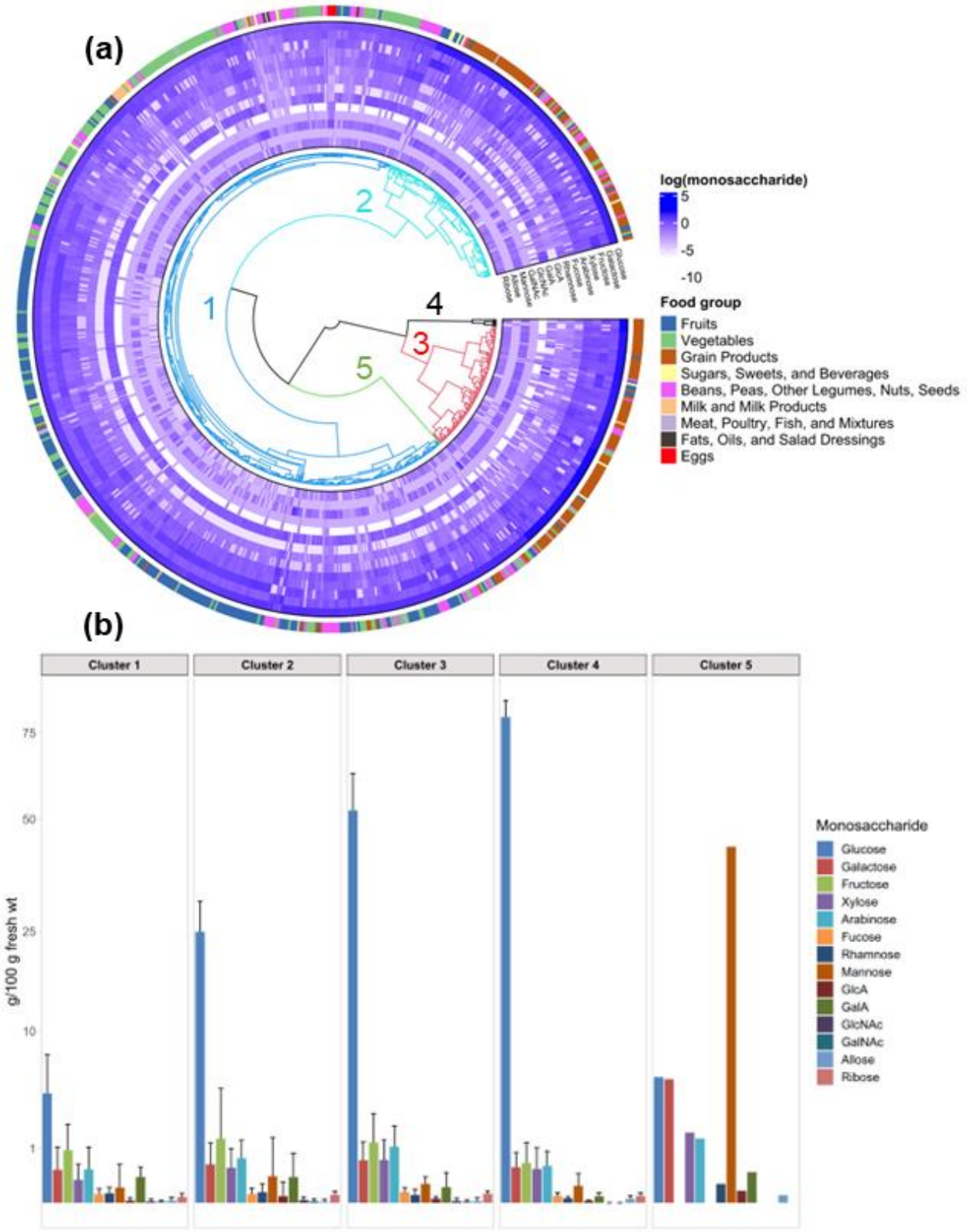


Figure 4.7. Clustering analysis of the Davis Food Glycopedia. (**a**) Circular heatmap and dendrogram of food samples based on their total monosaccharide composition profiles. Heatmap values are log-transformed. Outermost heatmap track corresponds to assigned food group for

each sample. (**b**) Cluster-averaged absolute monosaccharide compositions. Cluster numbers are indicated in (a) as shown.

A subset of the foods used in the Glycopedia was additionally analyzed using linkage and polysaccharide (FITDOG) analyses (**Figure 4.8**). As expected, grain products that are higher in glucose contained mainly starch corresponding to >80% of the polysaccharides based on FITDOG analysis (**Figure 4.8A**). The data were further corroborated with the abundance of 4-glucose in the linkage analysis (**Figure 4.8B**). Fruits and vegetables were found to contain diverse carbohydrate profiles consisting of the monosaccharides glucose, galactose, xylose, arabinose, and mannose (**Figure 4.8C**). The linkage and FITDOG analysis identified the polysaccharide structures as cellulose, xyloglucan, galactan, arabinoxylan, and mannans, respectively. Furthermore, fine variations in the linkage profile evince the presence of fine structures in specific polysaccharides. For example, the arabinan found in beans and peas is a linear structure comprised nearly exclusively of 5-linked and terminal (*t*-) arabinose, while the arabinan found in fruits and vegetables is a branched structure containing 2- and 3-linked arabinose (**Figure 4.8B**). The level of information obtained from this protocol is therefore unprecedented in terms of structural depth while providing enhanced throughput.

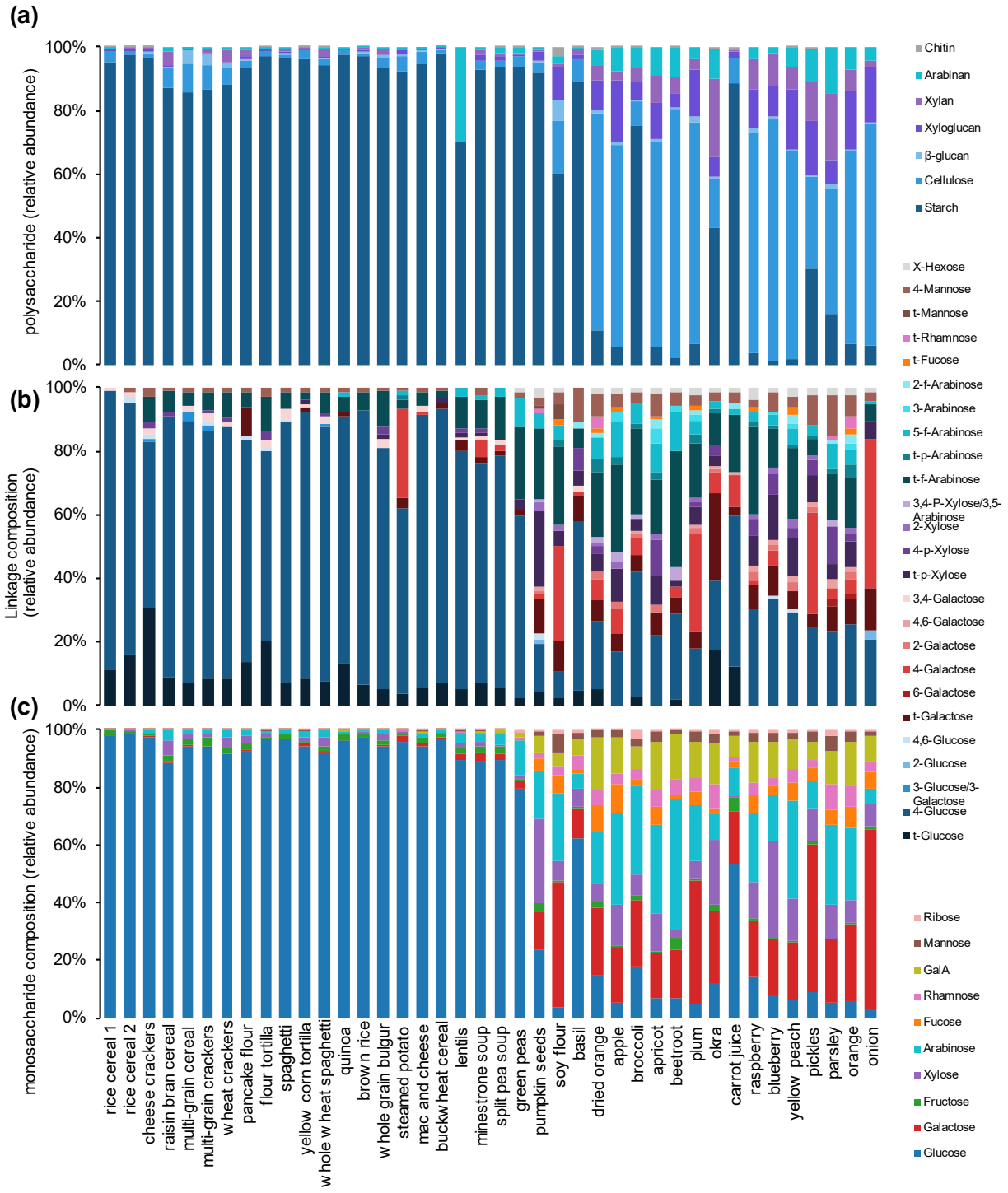


Figure 4.8. Multi-glycomics analysis of food. (a) Polysaccharide (FITDOG) composition, (b) glycosidic linkage composition), (c) monosaccharide composition.

References

1. 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.

2. Flint, H. J.; Scott, K. P.; Duncan, S. H.; Louis, P.; Forano, E., Microbial degradation of complex carbohydrates in the gut. *Gut Microbes* **2012,** *3* (4), 289-306.

3. Wardman, J. F.; Bains, R. K.; Rahfeld, P.; Withers, S. G., Carbohydrate-active enzymes (CAZymes) in the gut microbiome. *Nat Rev Microbiol* **2022,** *20* (9), 542-556.

4. Cronin, P.; Joyce, S. A.; O'Toole, P. W.; O'Connor, E. M., Dietary fibre modulates the gut microbiota. *Nutrients* **2021,** *13* (5).

5. Dhingra, D.; Michael, M.; Rajput, H.; Patil, R. T., Dietary fibre in foods: a review. *J Food Sci Technol* **2012,** *49* (3), 255-66.

6. Han, S.; Van Treuren, W.; Fischer, C. R.; Merrill, B. D.; DeFelice, B. C.; Sanchez, J. M.; Higginbottom, S. K.; Guthrie, L.; Fall, L. A.; Dodd, D.; Fischbach, M. A.; Sonnenburg, J. L., A metabolomics pipeline for the mechanistic interrogation of the gut microbiome. *Nature* **2021,** *595* (7867), 415-420.

7. 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.; Biddinger, S. B.; Dutton, R. J.; Turnbaugh, P. J., Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **2014,** *505* (7484), 559-63.

8. Koh, A.; De Vadder, F.; Kovatcheva-Datchary, P.; Backhed, F., From dietary fiber to host physiology: short-chain fatty acids as key bacterial metabolites. *Cell* **2016,** *165* (6), 1332-1345.

9. Morrison, D. J.; Preston, T., Formation of short chain fatty acids by the gut microbiota and their impact on human metabolism. *Gut Microbes* **2016,** *7* (3), 189-200.

10. Fischbach, M. A.; Sonnenburg, J. L., Eating For Two: How metabolism establishes interspecies interactions in the gut. *Cell Host Microbe* **2011,** *10* (4), 336-347.

11. Spencer, C. N.; McQuade, J. L.; Gopalakrishnan, V.; McCulloch, J. A.; Vetizou, M.; Cogdill, A. P.; Khan, M. A. W.; Zhang, X.; White, M. G.; Peterson, C. B.; Wong, M. C.; Morad, G.; Rodgers, T.; Badger, J. H.; Helmink, B. A.; Andrews, M. C.; Rodrigues, R. R.; Morgun, A.; Kim, Y. S.; Roszik, J.; Hoffman, K. L.; Zheng, J.; Zhou, Y.; Medik, Y. B.; Kahn, L. M.; Johnson, S.; Hudgens, C. W.; Wani, K.; Gaudreau, P. O.; Harris, A. L.; Jamal, M. A.; Baruch, E. N.; Perez-Guijarro, E.; Day, C. P.; Merlino, G.; Pazdrak, B.; Lochmann, B. S.; Szczepaniak-Sloane, R. A.; Arora, R.; Anderson, J.; Zobniw, C. M.; Posada, E.; Sirmans, E.; Simon, J.; Haydu, L. E.; Burton, E. M.; Wang, L.; Dang, M.; Clise-Dwyer, K.; Schneider, S.; Chapman, T.; Anang, N. A. S.; Duncan, S.; Toker, J.; Malke, J. C.; Glitza, I. C.; Amaria, R. N.; Tawbi, H. A.; Diab, A.; Wong, M. K.; Patel, S. P.; Woodman, S. E.; Davies, M. A.; Ross, M. I.; Gershenwald, J. E.; Lee, J. E.; Hwu, P.; Jensen, V.; Samuels, Y.; Straussman, R.; Ajami, N. J.; Nelson, K. C.; Nezi, L.; Petrosino, J. F.; Futreal, P. A.; Lazar, A. J.; Hu, J.; Jenq, R. R.; Tetzlaff, M. T.; Yan, Y.; Garrett, W. S.; Huttenhower, C.; Sharma, P.; Watowich, S. S.; Allison, J. P.; Cohen, L.; Trinchieri, G.; Daniel, C. R.; Wargo, J. A., Dietary fiber and probiotics influence the gut microbiome and melanoma immunotherapy response. *Science* **2021,** *374* (6575), 1632-1640.

12. 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).

13. Delannoy-Bruno, O.; Desai, C.; Castillo, J. J.; Couture, G.; Barve, R. A.; Lombard, V.; Henrissat, B.; Cheng, J.; Han, N.; Hayashi, D. K.; Meynier, A.; Vinoy, S.; Lebrilla, C. B.; Marion, S.; Heath, A. C.; Barratt, M. J.; Gordon, J. I., An approach for evaluating the effects of dietary fiber polysaccharides on the human gut microbiome and plasma proteome. *Proc Natl Acad Sci U S A* **2022,** *119* (20), e2123411119.

14. 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.; 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.

15. O'Grady, J.; O'Connor, E. M.; Shanahan, F., Review article: dietary fibre in the era of microbiome science. *Aliment Pharmacol Ther* **2019,** *49* (5), 506-515.

16. Gill, S. K.; Rossi, M.; Bajka, B.; Whelan, K., Dietary fibre in gastrointestinal health and disease. *Nat Rev Gastroenterol Hepatol* **2021,** *18* (2), 101-116.

17. 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.

18. Amicucci, M. J.; Nandita, E.; Lebrilla, C. B., Function without structures: The need for in-depth analysis of dietary carbohydrates. *J Agr Food Chem* **2019,** *67* (16), 4418-4424.

19. Xu, G. 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.

20. 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.

21. 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.

22. Galermo, A. G.; Nandita, E.; Barboza, M.; Arnicucci, 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.

23. 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.

24. Nandita, E.; Bacalzo, N. P.; Ranque, C. L.; Amicucci, M. J.; Galermo, A.; Lebrilla, C. B., Polysaccharide identification through oligosaccharide fingerprinting. *Carbohyd Polym* **2021,** *257*.

25. Wong, M.; Xu, G. G.; Park, D.; Barboza, M.; Lebrilla, C. B., Intact glycosphingolipidomic analysis of the cell membrane during differentiation yields extensive glycan and lipid changes. *Sci Rep-Uk* **2018,** *8*.

26. Park, D. D.; Xu, G.; Wong, M.; Phoomak, C.; Liu, M. Q.; Haigh, N. E.; Wongkham, S.; Yang, P. Y.; Maverakis, E.; Lebrilla, C. B., Membrane glycomics reveal heterogeneity and quantitative distribution of cell surface sialylation. *Chem Sci* **2018,** *9* (29), 6271-6285.

27. Chu, C. S.; Ninonuevo, M. R.; Clowers, B. H.; Perkins, P. D.; An, H. J.; Yin, H. F.; Killeen, K.; Miyamoto, S.; Grimm, R.; Lebrilla, C. B., Profile of native N-linked glycan structures from human serum using high performance liquid chromatography on a microfluidic chip and time-of-flight mass spectrometry. *Proteomics* **2009,** *9* (7), 1939-1951.

28. Barboza, M.; Pinzon, J.; Wickramasinghe, S.; Froehlich, J. W.; Moeller, I.; Smilowitz, J. T.; Ruhaak, L. R.; Huang, J.; Lonnerdal, B.; German, J. B.; Medrano, J. F.; Weimer, B. C.; Lebrilla, C. B., Glycosylation of human milk lactoferrin exhibits dynamic changes during early lactation enhancing its role in pathogenic bacteria-host interactions. *Mol Cell Proteomics* **2012,** *11* (6), M111 015248.

29. 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), 7471-80.

30. 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), 4138-51.

31. Li, Q. Y.; Xie, Y. X.; Wong, M. R.; Barboza, M.; Lebrilla, C. B., Comprehensive structural glycomic characterization of the glycocalyxes of cells and tissues. *Nature Protocols* **2020,** *15* (8), 2668-2704.

32. Pettolino, F. A.; Walsh, C.; Fincher, G. B.; Bacic, A., Determining the polysaccharide composition of plant cell walls. *Nat Protoc* **2012,** *7* (9), 1590-607.

33. Blakeney, A. B.; Harris, P. J.; Henry, R. J.; Stone, B. A., A simple and rapid preparation of alditol acetates for monosaccharide analysis. *Carbohyd Res* **1983,** *113* (2), 291-299.

34. Doares, S. H.; Albersheim, P.; Darvill, A. G., An improved method for the preparation of standards for glycosyl-linkage analysis of complex carbohydrates. *Carbohyd Res* **1991,** *210*, 311-317.

35. Anumula, K. R.; Taylor, P. B., A comprehensive procedure for preparation of partially methylated alditol acetates from glycoprotein carbohydrates. *Anal Biochem* **1992,** *203* (1), 101-108.

36. Rohrer, J. S., High-performance anion-exchange chromatography with pulsed amperometric detection for the determination of oligosaccharides in foods and agricultural products. *Acs Sym Ser* **2003,** *849*, 16-31.

37. Hanko, V. P.; Rohrer, J. S., Determination of carbohydrates, sugar alcohols, and glycols in cell cultures and fermentation broths using high-performance anion-exchange chromatography with pulsed amperometric detection. *Anal Biochem* **2000,** *283* (2), 192-199.

38. Castillo, J. J.; Couture, G.; Bacalzo, N. P., Jr.; Chen, Y.; Chin, E. L.; Blecksmith, S. E.; Bouzid, Y. Y.; Vainberg, Y.; Masarweh, C.; Zhou, Q.; Smilowitz, J. T.; German, J. B.; Mills, D. A.; Lemay, D. G.; Lebrilla, C. B., The development of the Davis Food Glycopedia-A glycan encyclopedia of food. *Nutrients* **2022,** *14* (8).

39. 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. *J Am Soc Mass Spectrom* **2021,** *32* (8), 2175-2185.

40. Nandita, E.; Bacalzo, N. P., Jr.; Ranque, C. L.; Amicucci, M. J.; Galermo, A.; Lebrilla, C. B., Polysaccharide identification through oligosaccharide fingerprinting. *Carbohydr Polym* **2021,** *257*, 117570.

41. Ndukwe, I. E.; Black, I.; Heiss, C.; Azadi, P., Evaluating the utility of permethylated polysaccharide solution NMR data for characterization of insoluble plant cell wall polysaccharides. *Anal Chem* **2020,** *92* (19), 13221-13228.

42. Perez Garcia, M.; Zhang, Y.; Hayes, J.; Salazar, A.; Zabotina, O. A.; Hong, M., Structure and interactions of plant cell-wall polysaccharides by two- and three-dimensional magic-angle-spinning solid-state NMR. *Biochemistry* **2011,** *50* (6), 989-1000.

43. Zhao, W. C.; Fernando, L. D.; Kirui, A.; Deligey, F.; Wang, T., Solid-state NMR of plant and fungal cell walls: A critical review. *Solid State Nucl Mag* **2020,** *107*.

44. Wu, Z. Q.; Serie, D.; Xu, G. G.; Zou, J., PB-Net: Automatic peak integration by sequential deep learning for multiple reaction monitoring. *J Proteomics* **2020,** *223*.

45. Ranque, C. L.; Stroble, C.; Amicucci, M. J.; Tu, D.; Diana, A.; Rahmannia, S.; Suryanto, A. H.; Gibson, R. S.; Sheng, Y.; Tena, J.; Houghton, L. A.; Lebrilla, C. B., Examination of carbohydrate products in feces reveals potential biomarkers distinguishing exclusive and nonexclusive breastfeeding practices in infants. *J Nutr* **2020,** *150* (5), 1051-1057.

46. 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.

Chapter 5. Summary and conclusions

While carbohydrates constitute a major portion of our usual diets, they have remained under-characterized because of the difficulties with carbohydrate analysis. Carbohydrates comprise a myriad of structures and this complexity and diversity of chemical structures make the analysis challenging. Recent advances in sample preparation, liquid chromatography, and mass spectrometry have enabled the development of glycomics method for the analysis of various forms of carbohydrates. Polysaccharide analysis using LC-MS/MS has been very limited and relied mostly on using enzymes and partial acid hydrolysis. We have developed a non-enzymatic, universal, reproducible chemical reaction to depolymerize polysaccharides into oligosaccharides using oxygen radicals. The FITDOG reaction enabled the LC-MS/MS analysis of oligosaccharides and thereby inferring polysaccharide structures based on fingerprint library. This workflow was further improved and expanded to be quantitative using external calibration curve. The quantitative FITDOG workflow can simultaneously identify and quantify up to 11 polysaccharide structures commonly found in food and plant samples. The information obtained from this single method is unprecedented in terms of coverage, multiplexing, quantitation, and throughput.

Integrating other LC-MS/MS-based glycomics workflow with the FITDOG method will provide the most comprehensive structural and quantitative information on food carbohydrates. Monosaccharide analysis can provide a very accurate and precise quantitation of the food total carbohydrates, or the ethanol-precipitated polysaccharides. Linkage analysis can give more structural insights based on the relative abundances of the glycosidic linkages in the sample. The FITDOG analysis can provide both structural and quantitative information, complementing both

monosaccharide and linkage data. These glycomics workflows were developed to be done with 96-well plates, increasing the throughput and being amenable to some automated liquid handling systems. LC-MS/MS provides the sensitivity and specificity needed to resolve closely related structures.

We envision that, with these recent advances in the field of glycomics analysis, researchers from different fields will be able to understand how carbohydrate structures affect functions. By doing more research and clinical studies that include in-depth and proper characterization of carbohydrates, we can improve the quality of the knowledge by generating structurally defined associations. These correlations can then generate more hypotheses to be tested, thereby increasing our understanding of the relation between carbohydrates and health. This understanding includes nutritional needs based on individual factors to aid in control of disease.