Unlocking Nature's Potential: Harnessing Omics Technologies for Sustainable Extraction of Functional Ingredients from Plants and Macroalgae

By

TRUC PHAM
THESIS

Submitted in partial satisfaction of the requirements for the degree of

MASTER OF SCIENCE

in

Food Science and Technology

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Daniela Barile, Chair

Juliana Maria Leite Nobrega de Moura Bell

Selina Wang

Committee in Charge

2023

Acknowledgements

I would like to express my deepest gratitude and appreciation to all those who have supported me throughout this journey and contributed to the success of this research. This journey has been a challenging and fulfilling one, and I could not have reached this milestone without the support and encouragement from the following individuals:

My advisor, Professor Daniela Barile, for her invaluable guidance, support, and admirable dedication to the success of all of her lab members. Her encouragement and guidance have been instrumental in shaping the direction of this thesis. Thank you Dani, for encouraging the direction of my research to grow in an area that I am deeply passionate about.

The Barile lab members, Randall, Sierra, Gulustan, Yu-Ping, Bruna, Mara, Oswaldo, Aidong, Ece, Han, and Yu, for your friendship and support. Thank you for being there with me every step of the way. I have learned so much from each of you, and you have been invaluable to my personal and academic growth.

The members of my thesis committee, Dr. Juliana Maria Leite Nobrega de Moura Bell and Dr. Selina Wang, for their commitment to serve as mentors. Their thoughtful comments have significantly enriched the quality of this work.

My collaborators for your contributions to this work. I especially want to thank Dr. Fernanda Furlan Goncalves Dias, Jasmin Yang, Mia Gaiero, and Andrea Tam from the Bell lab. My research would not have been possible without their input. These collaborations have enriched my graduate school experience and strengthened my skills as a researcher.

Finally, my family and friends for their love and support throughout my academic journey. Their belief in me kept me motivated to push forward and persevere. I especially want to thank my fiancé Cory Wake for his unwavering support. Cory, thank you for encouraging me to pursue my passions, feeding me delicious meals, and making me laugh even through some of the most stressful periods of my academic career. I love you, and I can’t wait to see what the future brings for us.

Thank you all for being an integral part of this remarkable journey. Your support and contributions have truly made this thesis possible

Abstract

With growing environmental concerns, the need for sustainable practices in the food industry has become more pressing than ever before. While certain plant species such as soy have gained popularity as valuable food sources, there are numerous other underutilized plants, including beans, lentils, and marine organisms such as macroalgae, which remain largely unstudied. These overlooked plants offer not only nutritional benefits, such as high protein content, fiber, complex carbohydrates, and bioactive compounds, but are also key to reducing environmental impact through decreased land and water usage, and their crucial roles in sustainable ecosystems and agriculture. Despite these benefits, the incorporation of these underutilized plants in food products like snacks and plant-based meat alternatives have been limited, largely due to inadequate characterization of these plants. The lack of in-depth understanding of these plants' macromolecular compositions and the impact of green extraction processes has impeded their full utilization. This thesis serves as the first steps to addressing this gap by employing proteomics and glycomics techniques to evaluate the effects of green processing strategies on the release of valuable macromolecules such as proteins and carbohydrates. By understanding the influence of environmentally friendly extraction methods on plant protein and carbohydrate compositions, this work serves to guide the development of sustainable food processing strategies with the aim of increasing the diversity of sustainable plants utilized in the food industry. Ultimately, harnessing the full potential of underutilized plants can contribute to mitigating environmental issues and promote a more resilient and sustainable food ecosystem.

Chapter I introduces proteomics and glycomics as strategies to increase understanding of proteins and carbohydrates of underutilized plants and the influence of green processing techniques on these macromolecules. The environmentally friendly extraction processes featured in this work are also discussed.

Chapter II focuses on the proteomics and glycomics analyses of black beans (*Phaseolus vulgaris*). The effects of green extraction methods on protein composition are investigated, and oligosaccharides stemming from the breakdown of bean waste byproducts are characterized.

Chapter III utilizes proteomics and glycomics techniques to characterize the composition of *Lens culinaris Medik*, commonly known as lentils. The impact of green extraction methods on protein composition and the characterization of oligosaccharides derived from the breakdown of lentil waste byproducts are investigated.

Chapter IV examines the proteins and carbohydrates extracted from giant kelp (*Macrocystis pyrifera*), a macroalgae from the Pacific Ocean. Proteins, oligosaccharides, and monosaccharides are characterized to guide the development of environmentally friendly food processing technologies.

Table of Contents

Acknowledgements ii
Abstract iv
Table of contents vi
CHAPTER 1: 1
Background - Proteomics 2
Plant proteomes 2
Proteomics sample preparation and methods of analysis 3
Proteomics and food science 7
Application of proteomics in the presented work 8
Background - Glycomics 8
Plant glycomes 9
Plant glycomics sample prep and analysis 11
Glycomics and food science 14
Application of glycomics in the presented work 15
Background - Green extraction processes 16
Aqueous extraction, enzyme-assisted aqueous extraction, and Osborne fractionation 16
Subcritical water extraction 18
Microwave extraction 19
Application of green extraction technologies in the presented work 20
References 21
CHAPTER 2: 28
Abstract 29
Introduction 30
Bean composition 31
Bean proteins 31
Bean carbohydrates 33
Protein and carbohydrate analysis in the presented work with black beans 34
Experimental procedure 35
Protein extraction methods 36
SDS-PAGE 37
Proteomics of black bean extracts 38
Glycoprofiling of subcritical extracts 40
Results 43
Total protein extracted 43
SDS-PAGE 44
Proteomics results of AEP, EAEP, and Osborne fractions 45
Glycoprofiling of subcritical water extracts 51
Discussion 61
Bean proteins (AEP, EAEP, Osborne fractionation) 61
Bean insoluble subcritical water extracts glycoprofiling 63
Conclusion 66
Acknowledgements 68
Supplementary materials 69
References 109
CHAPTER 3: 115
Abstract 116
Introduction 117
Lentil composition 118

Lentil proteins 119
Lentil carbohydrates 120
Protein and carbohydrate analysis in the presented work with lentils 123
Experimental procedure 124
Extraction methods 125
SDS-PAGE 126
Proteomics of Osborne fractions, AEP, EAEP 126
Glycoprofiling of subcritical samples 129
Results 131
Total protein extracted 131
SDS-PAGE 133
Proteomics results of AEP, EAEP, and Osborne fractions 134
Glycoprofiling of subcritical water extractions 141
Discussion 153
Lentil proteins (AEP, EAEP, Osborne fractionation) 153
Lentil insoluble subcritical water extraction glycoprofiling 156
Conclusion 159
Acknowledgements 160
Supplementary materials 161
References 182
CHAPTER 4: 189
Abstract 190
Introduction 191
Algae composition 193
Algae proteins 194
Algae carbohydrates 195
Protein and carbohydrate analysis in the presented work with algae 196
Experimental procedure 197
Extraction methods 197
Proteomics of algae extracts 198
Oligosaccharide profiling of algae extracts 200
Monosaccharide analysis of algae extracts 201
Results 203
Selected extraction conditions for functional compounds 203
Proteomics results 204
Glycoprofiling results 205
Monosaccharide quantification 212
Bioactivity results (antidiabetic activity) 219
Discussion 220
Algae proteins 220
Algae carbohydrates 223
Conclusion 226
Acknowledgements 227
Supplementary materials 228
References 235

CHAPTER 1:

Proteomics and glycomics to characterize plants and guide the development of green extraction techniques

Background - Proteomics

Proteomics, or the analysis of proteomes, has evolved significantly over the decades. Broadly speaking, proteomics combines two major techniques: separation and characterization. Presently, it is the only technique capable of simultaneously resolving hundreds of proteins in one procedure. The term “proteome” was first coined in the early 1990s, and the initial definition of the proteome, the "protein complement to the genome", has expanded as advancements in analytical instruments have allowed for more comprehensive investigations of protein expression. Specifically, these advanced analytical instruments allow exploration into a broader range of expressed proteins, as well as a wider range of analyte concentrations that can be detected simultaneously in a given sample (Wilkins et al., 1996). Now, proteomics encompasses not only the identification of proteins but also their quantification, the identification of posttranslational modifications (PTMs), the investigation of protein-protein interactions, and protein profiling (Baginsky, 2009). Protein profiling aims to comprehensively identify proteins which make up the major constituents of a proteome of interest. Broadly speaking, proteomics is the study of protein structure, either in their intact form or after proteolysis.

Plant proteomes

The complexity of proteomes is particularly reflected in plants, where protein expression is regulated based on the plant's life cycle stage or specific plant parts. Proteomes within a plant exhibit variations across different plant parts and organelles (Hajduch et al., 2007). Plant proteomics encompasses the analysis of plants, organs, tissues, cells, and organelles both within and outside plant cells (Šamaj and Thelen, 2007). In general, protein functions are dictated by physicochemical properties stemming from the structure of the protein. Understanding the relationship between protein structure and function is vital to realizing the full protein potential of the underutilized plants featured in this work (C.-H. Tang and Sun, 2011), particularly in the context of enhancing their utilization in food products.

To date, plant proteomics has primarily focused on elucidating the molecular mechanisms and functional interactions of plant proteins involved in essential processes such as cell signaling, protein turnover, and membrane association (Chen and Harmon, 2006; Šamaj and Thelen, 2007). With this focus, plant proteomics has been employed to investigate plant and crop production, growth, development, stress, climate change effects, and interaction between plants and microorganisms (Chen and Harmon, 2006; Jorrin-Novo et al., 2020).

Proteomics sample preparation and methods of analysis

Proteomics analysis can be divided into two parts: separation of protein/peptides, and identification and characterization of proteins/peptides, often achieved through mass spectrometry (MS) (Šamaj and Thelen, 2007). While two-dimensional gel electrophoresis, a technique in which proteins were separated in based on their isoelectric points and molecular weights, has been considered the cornerstone of plant proteomics, significant advancements in MS technologies and the development of user-friendly MS interfaces and mass spectral search algorithms have increasingly made MS the primary tool for protein identification and characterization (Chen and Harmon, 2006; Šamaj and Thelen, 2007). Analysis by MS requires ionization and volatilization of the analyte into the gas phase. In early MS technologies, this requirement made studying proteins and peptides challenging, but the development of soft ionization techniques like matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI) have enabled MS-based proteomics. Today, the most commonly used MS sources for proteomics are MALDI and ESI, which allow ionization of larger and traditionally difficult to ionize molecules without unwanted in-source fragmentation. For instances in which prior separation is desirable, the use of low-throughput techniques such as electrophoresis gels has now been replaced by reverse-phase liquid chromatography coupled online with mass spectrometry. The mass detectors used range from the simpler time-of-flight (TOF) analyzers to

more complex devices such as Fourier Transform ion-cyclotron resonance (FT-ICR) analyzers. For analysis by MALDI-ToF, proteins are mixed with a matrix able to absorb the laser's energy, dried to a crystalline state onto a stainless steel plate, and subjected to a laser under vacuum. This process causes vaporization and ionization of both the matrix and analyte, and a signal is recorded by the ToF to determine their mass-to-charge ratio (m/z), generating mass spectra (Q. Zhang, 2017). On the other hand, ESI involves the volatilization and ionization of proteins or peptides dissolved in a liquid solution. To achieve this, a solution containing the analyte(s) is passed through a high potential capillary, creating an electrostatic spray and charged droplets. The charged droplets are evaporated to reduce droplet size, and upon sufficient desolvation, Coulomb repulsion forces causes the droplet to explode, releasing the charged analyte(s) into the gas phase for MS analysis (Šamaj and Thelen, 2007; Q. Zhang, 2017). The advancement of MS technologies in combination with user-friendly software for mass spectral data searching has enabled high-throughput proteomics analysis and interpretation (Šamaj and Thelen, 2007).

MS-based proteomics can be divided into two approaches: top-down and bottom-up. Top-down proteomics analyzes proteins in their intact form, allowing direct measurement of protein molecular weight and straightforward quantification of purified protein samples. Tandem mass spectrometry is employed to obtain sequence data through controlled fragmentation, enabling the identification of the intact protein, isoforms, and post-translational modifications (PTMs) such as phosphorylation and glycosylation (Toby et al., 2019). However, top-down approaches may be limited by the ability to ionize and fragment large proteins in complex samples. In contrast, bottom-up proteomics analyzes peptides. Bottom-up may also be referred to as shotgun proteomics, which have similar workflows, but the term "bottom-up" usually refers to the analysis of purified proteins, while "shotgun proteomics" refers to analysis of a mixture of proteins (Yates, 2004). Bottom-up and shotgun proteomics utilizes specific enzymes with known cleavage sites to proteolytically digest proteins and analyze the resulting peptides. Mass spectrometry is used

to obtain sequence data of the peptides, which are then searched against databases to identify proteins. Ionization and fragmentation of peptides in bottom-up approaches are generally less problematic than ionization of intact proteins in top-down approaches (Y. Zhang et al., 2013).

Bottom-up and shotgun approaches require enzymatic digestion of proteins prior to analysis, allowing for the analysis of complex protein mixtures and thousands of peptides in one experiment. These approaches offer higher throughput and minimize sampling errors associated with top-down approaches, particularly for proteins with extreme molecular weights, low abundance, or high hydrophobicity (Chen and Harmon, 2006). Bottom-up and shotgun approaches do not require any prior knowledge of proteins to be analyzed, making them well suited for discovery studies or when working with novel or understudied matrices. Particularly for new studies where there is not a lot of previously generated data, MS/MS spectral searching is highly useful for the mentioned reasons.

However, bottom-up and shotgun approaches also have limitations. These approaches heavily rely on protein databases for protein identification. This reliance on databases to obtain rich protein identification data can be a disadvantage for bottom-up approaches, particularly for understudied matrices. Protein identification using databases is usually best for organisms with complete genome sequence data (Chen and Harmon, 2006). Additionally, the accuracy of protein identification through database searching is not as high as that of *de novo* peptide sequencing, especially for plant species lacking genetic resources for database comparisons (Šamaj and Thelen, 2007).

Plant proteomics sample preparation and analysis present numerous challenges. Sample preparation methods for proteomics are inherently selective and unable to analyze all protein components simultaneously due to sample complexity, the diverse chemical nature of proteins,

or instrument detection limitations (Chen and Harmon, 2006). This challenge becomes prominent when analyzing protein mixtures extracted by different solvent conditions. Any one method of proteomics analysis can only analyze a subset of the present proteins depending on the efficiency of their extraction, precipitation, or solubilization (Agrawal et al., 2013). In the context of plant proteomics, high carbohydrate-to-protein ratio in most plant tissues and extracts may hinder the isolation and analysis of proteins (Nagai et al., 2008; Šamaj and Thelen, 2007). Moreover, with a bottom-up proteomics approach which relies on database searching to identify proteins, the availability (or lack thereof) of species-specific protein databases greatly influences the quality of the generated protein data.

Figure 1.1 presents a typical bottom-up proteomics workflow. The process begins with protein extraction, precipitation to concentrate proteins and remove interfering compounds, solubilization of the protein pellet, reduction, alkylation, enzymatic hydrolysis (typically using trypsin), and solid-phase extraction (SPE) to prepare samples for liquid chromatography-mass spectrometry (LC-MS/MS) analysis. Upon injection onto an LC/MS instrument, peptides are separated by liquid chromatography (LC), and peptide sequence data is acquired by MS and tandem MS. During MS analysis, a precursor peptide is selected, then fragmented to generate a tandem MS (MS/MS) product ion spectrum (Q. Zhang, 2017). The MS/MS spectra are then used to determine peptide sequence, either through manual interpretation (*de novo* sequencing) or by searching the spectra against a database to obtain a partial or full sequence match (Šamaj and Thelen, 2007). The precursor mass, partial sequencing obtained through tandem MS, and knowledge of enzymatic cleavage sites (typically trypsin is employed in proteomics analyses due to its highly specificity) are all utilized to search databases to identify proteins.

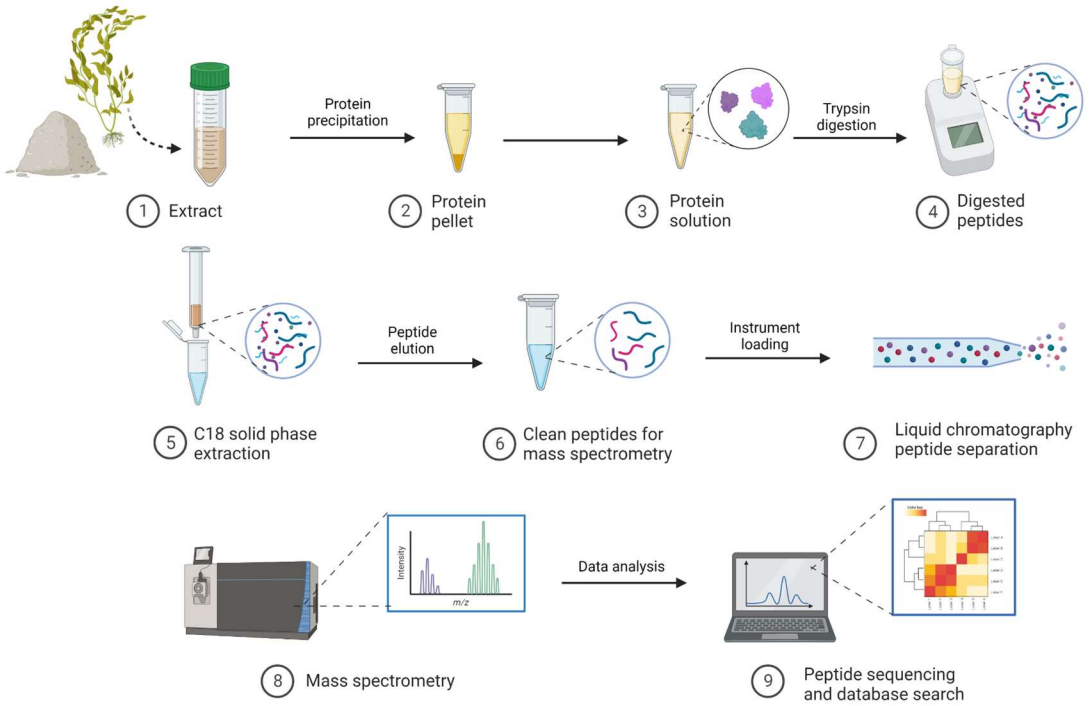


Figure 1.1. Overview of bottom-up proteomics workflow. Created with BioRender.com

Proteomics and food science

In the field of food science, proteomics and protein profiling are employed within three main topics: food quality, safety, and authentication; investigating food processing techniques and the effects of food processing; and characterization of food ingredients and their nutritional aspects (De Angelis et al., 2008; Gallardo et al., 2013; Huang et al., 2022; Pavlovic et al., 2013).

Food processing affects the “technological, sensorial, and nutritional qualities" of food (Gallardo et al., 2013). During processing, proteins undergo various non-enzymatic posttranslational modifications (PTMs) through oxidation reactions, the Maillard reaction, condensation, loss of side chains, and proteolytic degradation (Pischetsrieder and Baeuerlein, 2009). These modifications as a result of processing alter protein properties such as solubility and hydrophobicity (Gallardo et al., 2013; Pischetsrieder and Baeuerlein, 2009), emphasizing the

need for comprehensive proteome characterization to assess the impact of processing and proteins on food products.

In the case of plant proteins, the presence of rigid cell walls poses a challenge for their extraction. Physical, chemical, or enzymatic treatments are typically required to disrupt plant cell walls and release proteins that are protected within (Šamaj and Thelen, 2007). Various extraction methods, either alone or in combination, are utilized to account for the diversity of protein solubilities and plant tissue compositions, ensuring effective protein extraction (Hurkman and Tanaka, 2007).

Application of proteomics in the presented work

In this study, proteomic analysis is employed to identify the specific proteins present in sequentially extracted Osborne fractions (albumin-rich, globulin-rich, prolamin-rich, and glutelin-rich fractions) as well as pulse proteins extracted through aqueous extraction processes (AEP) and enzyme-assisted extraction processes (EAEP). Additionally, proteomic techniques provide insights into how proteolysis affects the protein profile of the EAEP extract and enable the identification of hydrolyzed proteins. This research aims to shed light on the protein composition of black beans and lentils across different extraction methods, facilitating the development of efficient and sustainable approaches for obtaining high-quality protein ingredients deriving from beans and lentils.

Background - Glycomics

Glycomics is a growing field (after genomics and proteomics) focused on investigating the glycome, or the whole ensemble of sugars (free or bonded) within an organism, in order to understand the complex structure-function relationships of carbohydrates, also known as glycans (Yadav et al., 2015). Glycans are composed of sugar monomers linked by glycosidic

bonds. They serve essential metabolic, structural, and physical roles, and they are involved in many biological processes such as cell-cell communication, cell recognition, disease states such as cancer, and stress response (Mahajan et al., 2021; Varki, 2017; Yadav et al., 2015). The function of glycans is often dictated by their specific structures, thus obtaining structural information is vital to fully understanding their diverse roles.

Glycans can be categorized by structure or by constituent monosaccharides. In terms of structure, they can be linear or branched (Lannoo et al., 2014). Based on constituent monosaccharides, glycans may be classified as homopolysaccharide or heteropolysaccharide. Homopolysaccharides are composed of a single type of monosaccharide, such as plant cellulose which consists of linear chains of glucose units connected by β-1,4 linkages. In contrast, heteropolysaccharides are composed of two or more different monosaccharides or monosaccharide derivatives. Heteropolysaccharides also include glycoconjugates like glycoproteins or glycolipids (Yadav et al., 2015).

Plant glycomes

Food glycomes vary depending on the source, whether derived from plants, animals, or microbes, and each source exhibits different structures (W. Tang et al., 2022). Plant-derived sources, in particular, possess abundant glycans that are not found in animal sources, in part due to the presence of plant cell walls. Using plant glycomic approaches, glycome profiling is well-suited for screening of plant cell wall and biomass samples (Yadav et al., 2015).

Plant glycomes encompass various types of carbohydrates including mono-, oligo, and polysaccharides, all of which play important roles in food science and human health (An and Lebrilla, 2010; Bose et al., 2021). For example, oligosaccharides have demonstrated abilities to serve as a prebiotic supplement in foods, a method for fruit preservation, and as a fat replacer

(Bose et al., 2021; Illippangama et al., 2022; Peng and Yao, 2017). Glycans also exist in plants as independent entities or in conjugated forms known as glycoconjugates. In this form, sugars may be linked in non-carbohydrate molecules such as lipids and proteins. These glycoconjugates are respectively referred to as glycolipids and glycoproteins (Lannoo et al., 2014; Yadav et al., 2015). In general, carbohydrates exert substantial influence on food flavor, texture, and nutritional and functional activities (Peng and Yao, 2017), emphasizing the importance of understanding their structural characteristics for their effects as food ingredients.

In plants, glucose is the most common monosaccharide making up glycans. Other constituent monosaccharides include fructose, galactose, galactose, mannose, arabinose, fucose, and xylose. Monosaccharide derivatives, including simple sugar acids and amino sugars and their derivatives, are also present (Yadav et al., 2015).

Plant glycomes exhibit variations across different organelles. Select types of carbohydrates are present in specific organelles, such as cell walls, mitochondria, and chloroplasts (Mahajan et al., 2021). Cell walls, located on the outermost layer of plant cells, play critical roles in maintaining tissue and organ shape, facilitating interactions between intracellular and extracellular macromolecules, providing defense mechanisms and overall ensuring proper cell functioning (Keegstra, 2010). To serve in these diverse roles, cell walls are rigid for structural support while simultaneously being flexible and semipermeable. The three important polysaccharides in plant cell walls are cellulose, hemicellulose, and pectin (Mahajan et al., 2021; Voiniciuc et al., 2018).

Cellulose, the primary component of plant cell walls, is the most abundant polysaccharide on Earth (Mahajan et al., 2021). Its structure may vary depending on the specific plant part, but in general, it is composed of-glucose residues connected by β-(1,4)-linkages. The chains are intertwined, forming a complex network held together by hydrogen bonding (Mahajan et al.,

2021). Hemicellulose, another essential component of plant cell walls, is a heteropolysaccharide composed of pentose and hexose sugars. The pentoses include arabinose and xylose (and fucose, in the case of algae), and the hexoses include glucose, mannose, and galactose (Buckeridge et al., 2000; Schädel, Blöchl, Richter, and Hoch, 2010). Lastly, pectin, also referred to as pectic polysaccharide, consists primarily of D-galacturonic acid (Sriamornsak, 2003).

Plant glycomics sample prep and analysis

A number of tools have been employed in the field of glycomics including chromatography, mass spectrometry (ESI-MS, MALDI-MS), NMR spectroscopy, and carbohydrate microarrays (Mahajan et al., 2021; Yadav et al., 2015). Among these techniques, mass spectrometry is the most widely utilized, with MALDI-ToF-MS being the second most commonly used for generating glycan structure data (W. Tang et al., 2022). The expansion of rapid glycan analysis has been made possible, in part, by the well-established field of proteomics and the development of innovative MS technologies coupled with chromatography and bioinformatics (Goli et al., 2021; Yadav et al., 2015). Technological advancements such as electrospray ionization has allowed for the analysis of larger macromolecules that were previously challenging to volatilize (Q. Zhang, 2017). MS is highly selective, sensitive, and is able to rapidly scan samples during analysis (Yadav et al., 2015), making it an invaluable tool for studying complex mixtures and systems like glycomes.

Despite the advancements in MS technology for studying glycan structure, many challenges still remain in the field of glycomics. The structural complexity of glycans is unlike those of nucleic acids and proteins, and the diversity and complexity of glycan structures make structure identification inherently difficult. Glycans vary by constituent monosaccharides, types of linkages, and branching (Mahajan et al., 2021). The monosaccharides themselves are chemically similar and difficult to differentiate solely through MS. While structural differences in

glycans may lead to functional variations, distinguishing between them can be difficult, often requiring multiple complementary approaches to elucidate their complete structures. Further, while glycomics utilizes similar techniques and instruments to other omics fields, it has historically trailed behind because it lacks the predictive nature that fields like genomics and proteomics are able to capitalize on. The advancements made in genomics and proteomics have not directly translated to glycomics because sequence information of DNA and proteins is not linked to the biosynthesis of glycan chain composition, branching, and linkage (Yadav et al., 2015). Consequently, the progress made in other omics fields cannot be readily extended to the field of glycomics.

In addition to the challenges related to structural complexity, the analysis of glycans presents its own set of unique challenges. Glycan isolation can be challenging, particularly due to the lack of high-throughput techniques. Additionally, MS is limited in the type of data it is able to generate. While highly accurate mass data is a key aspect of MS analysis, it cannot differentiate between monosaccharides with identical mass but different spatial arrangements. Determining the exact constituent monosaccharides and stereochemistry of glycan linkages requires additional derivatization steps to be amenable to MS.

To differentiate features, MS relies on retention time differences arising from the liquid chromatography pre-separation component to discern isomers with the same mass and fragmentation pattern. However, ionization issues such as in-source fragmentation and dimerization can arise, necessitating the search and summation of multiple features to approximate the quantity of a single glycan (Huang et al., 2022). Unlike other omics fields, there is a general lack of user-friendly data analysis software specifically tailored to process glycan data, thus manual inspection and annotation of thousands of spectra is often required. Lastly, the synthesis of specific glycans in high purity and isomer selectivity is difficult, leading to a lack

of analytical standards and reference standards that may be used for identity confirmation. To overcome this limitation, enzymes with highly specific cleavage sites and liquid chromatography are often employed to deduce the complete structure of a glycan after additional fractionation/purification steps. However, these methods are time consuming and tedious (Ding et al., 2009; Patwa et al., 2009).

Figure 1.2 presents a typical glycomics workflow. For food matrix analysis, the process begins with pretreatment of the ingredient, commonly involving grinding or milling to facilitate glycan release. Extraction methods, which may include enzymatic hydrolysis, are employed to release glycans from their biological matrices. Protein isolation and removal is achieved through alcohol precipitation, while sugars remain in solution. To prepare samples for MS profiling, a two-stage solid-phase extraction (SPE) approach utilizing C18 and porous graphitized carbon (PGC) is utilized. Upon injection onto an LC/MS instrument, oligosaccharides are separated by liquid chromatography (LC), and compositional data is acquired by MS and tandem MS. Specifically, this work employs the use of nano-LC quadrupole time-of-flight (QToF) instrument. Partial structural information can be obtained through LC separation based on retention time, oligosaccharide mass as determined by MS, and constituent monosaccharides as determined by fragmentation data. Lastly, data analysis entails meticulous inspection, interpretation, and annotation of fragmentation data, in order to create novel libraries for sample identification. This process involves manual inspection and annotation of thousands of spectra, ultimately resulting in the development of a unique database for each project.

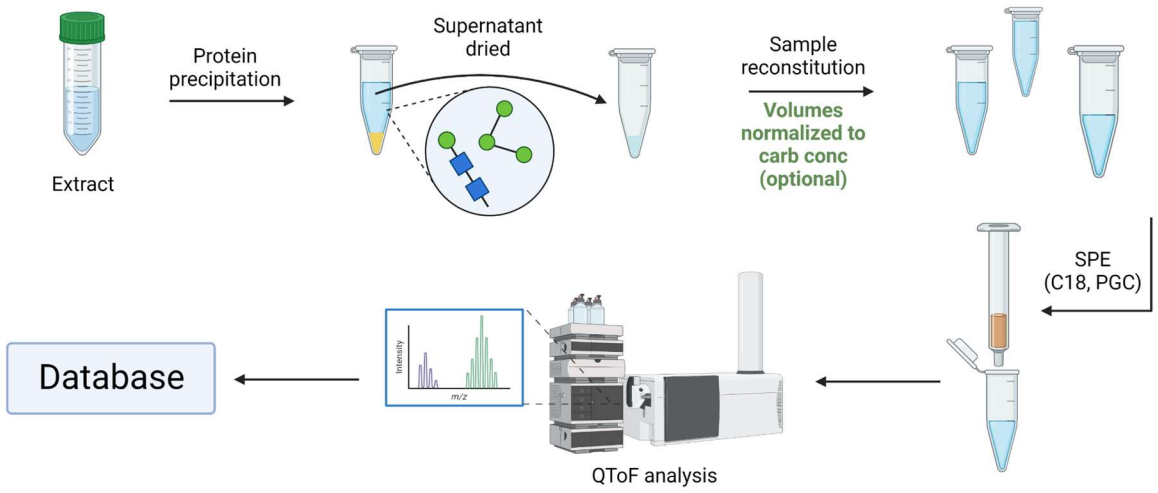


Figure 1.2. Overview of general glycomics workflow. Created with BioRender.com

Glycomics and food science

Carbohydrates play a significant role in shaping food flavor and texture, making them important targets for controlling food quality and sensory attributes (Serin and Sayar, 2016). In the context of the food matrix, glycomics can be used to evaluate food safety, quality, and authenticity (Nandita et al., 2021). Further, from a functional foods and infant-nutrition perspective, glycomics has found widespread application in the study of milk, focusing on human milk oligosaccharides (HMOs) and bovine milk oligosaccharides (BMOs) which play key roles in human health and development (De Leoz et al., 2015; Zivkovic and Barile, 2011). In particular, food glycomes intricately interact with gut microbiota. Glycans can act as prebiotics, resisting digestion and undergoing fermentation in the large intestine, thereby influencing the gut microbiota and impacting immune function (Barile and Rastall, 2013; Vulevic et al., 2008). Additionally, glycans can provide protection against pathogens and support the development of a healthy microbiome, particularly in infants (De Leoz et al., 2015).

In the context of food processing, the application of glycomics techniques for evaluating food extraction technologies remains limited. Glycome profiling of legumes primarily include studies

of cell wall structure of various plant parts and plant stress physiology (Muszyński et al., 2015; Singh et al., 2021; Zadražnik et al., 2017). More broadly in the field of foodomics, glycomics has been utilized to improve agriculture by monitoring crop health and fruit ripening, improving utilization of the entire plant, and gaining insights into crop production under environmental stress conditions (Couture et al., 2021; Oliveira et al., 2020; Singh et al., 2021; Zhao et al., 2023). Furthermore, glycomics has been utilized in the analysis of food matrices to understand the relationship between organoleptic qualities, molecular structure, food quality (Moreira et al., 2015; Nandita et al., 2021).

Application of glycomics in the presented work

This work utilizes a discovery-based MS approach for glycome profiling, specifically focusing on oligosaccharide profiling of plant extracts intended for food applications. MS analysis is employed to determine oligosaccharide mass, and tandem MS is utilized to identify the constituent monosaccharides. The main objective of this work is to characterize the composition of free oligosaccharides and monosaccharides released through various extraction techniques including enzymatic or acidic hydrolysis and subcritical water extraction.

Glycomic techniques are employed to investigate the breakdown of the predominant waste product resulting from various extraction processes (known as insoluble fraction) by subcritical extraction, identify the release of potentially prebiotic compounds through subcritical extraction, and determine the compositional differences in insoluble fraction that stem from various eco-friendly extraction methods (aqueous, microwave-assisted, and enzyme-assisted extractions). By identifying glycan structures in the insoluble fraction resulting from various extraction conditions, we aim to gain insights into the effects of green extraction methods and conditions and provide guidance for the development of more effective and innovative extraction processes. Furthermore, by exploring these compounds, we seek to uncover additional value

and utility that can be derived from this extraction byproduct, contributing to a more sustainable and resource-efficient approach.

Background - Green extraction processes

Green extraction methods seek to reduce the time, energy, and toxicity involved with traditional extraction methods. Traditional extraction methods (e.g. Soxhlet, liquid-liquid) requires extended extraction times, sizable amounts of samples, and large volumes of flammable and hazardous organic solvents. These extraction methods result in a substantial amount of generated solvent waste and environmental complications (Mendiola et al., 2013). Green extraction approaches avoid the use of hazardous organic solvents and aim to selectively extract and isolate bioactive compounds using environmentally friendly methods (J. Zhang et al., 2020). These approaches capitalize on properties such as polarity, temperature, and pressure to achieve faster extractions with optimization opportunities. Techniques such as ultrasound-assisted extraction, microwave-assisted extraction, and subcritical water extraction are examples of green extraction methods (Mena-García et al., 2019) that have demonstrated numerous advantages when compared with traditional extraction approaches.

Aqueous extraction, enzyme-assisted aqueous extraction, and Osborne fractionation

Current commercial methods for extracting protein from pulse flours involve alkaline extraction at pH 8-11 (Boye et al., 2010). These aqueous extraction processes (AEP) offer advantages in terms of scalability, high extraction and recovery yields, and cost-effectiveness (Cheng et al., 2019; de Moura et al., 2009). However, limited information is available regarding the influence of extraction conditions on bean protein extractability, physicochemical properties, and functionality. To our knowledge, only one study thus far has optimized the aqueous extraction of protein from pinto beans, but yielded relatively low total protein extractability (Tan et al., 2014). For lentils, there is slightly more literature exploring the aqueous extraction of lentil proteins

(Jarpa-Parra et al., 2014; Joshi et al., 2011; Lee et al., 2007), but this research is still lacking compared to other pulse crops such as soybean and chickpea. Therefore, further research is needed to develop effective extraction processes that produce pulse protein ingredients suitable for widespread commercial applications.

A useful modification of the AEP to improve protein extractability is the enzyme-assisted aqueous extraction process (EAEP), which involves the addition of proteases and/or carbohydrases enzymes to the extraction slurry. EAEP has found extensive application in the processing of oil-bearing materials like soy, rapeseed, and sunflower. This technique enhances the extraction of both oil and protein by disrupting cell walls and lipid body membranes, which are targeted by the actions of carbohydrases and proteases, respectively (Campbell et al., 2016; de Moura et al., 2008).

However, in the case of low-oil materials like pulses, the primary role of proteases is to hydrolyze proteins into smaller, more soluble subunits or peptides, thereby increasing cell matrix porosity (Rosenthal et al., 1998; Souza Almeida et al., 2021). This process allows previously entrapped proteins to be released at the surface of flour particles and extracted into the aqueous phase. Proteolysis not only improves extractability from the flour but also modifies protein structure, therefore influencing the functional and biological properties of the extracted proteins. However, studies focusing on common beans and lentils and their protein classes, as well as the impact of extraction conditions on protein properties and yield, remain limited.

In the development of scalable and sustainable aqueous extraction methods like AEP and EAEP, it is useful to include other extraction methods as a point of reference. The effective development of AEP and EAEP necessitates a comprehensive understanding of different protein classes found in common beans and how these classes are influenced by various

extraction conditions. To achieve this, Osborne fractionation can be employed to gain a general understanding of the properties and extractability of individual protein classes.

Osborne fractionation has been employed for nearly a century as a simple strategy to sequentially extract protein classes from plant material based on their solubility (Osborne, 1924). By employing various extraction media, protein classes with varying solubilities can be selectively extracted. Sequential extraction using water, dilute salt solution, water/alcohol solution, and dilute alkali solution yields the albumins, globulins, prolamins, and glutelins fractions, respectively. It should be noted that these fractions are also referred to as albumin-rich, globulin-rich, prolamin-rich, and glutelin-rich fractions, as no purification of proteins occurs between extractions, resulting in the presence of protein classes in multiple fractions (Yang et al., 2023). Although Osborne fractionation is not a scalable extraction method, it serves as a benchmark for the development of AEP and EAEP approaches.

Subcritical water extraction

Subcritical water extraction is an environmentally friendly technology used for valorizing byproducts via extraction from processing streams (J. Zhang et al., 2020). Due to water's molecular structure and considerable hydrogen bonding, it typically serves as a highly polar solvent with a high dielectric constant at ambient temperature and pressure. For this reason, it is not effective for extracting nonpolar compounds. Instead with subcritical water extraction, high temperature and high pressure are manipulated to affect the physicochemical properties of solvents (Zhang et al. 2020). This technique uses water as the sole solvent, capitalizing on water's unique properties while avoiding the use of toxic organic solvents.

Subcritical water uses water between 100 and 374 °C and sufficiently high pressure to maintain a liquid state (Haghighi and Khajenoori, 2013). Increased temperature leads to decreased

polarity, viscosity, surface tension, and increased diffusivity (J. Zhang et al., 2020). Under these conditions, the dielectric constant of water is comparable to that of organic solvents, allowing for the extraction of low polarity compounds. Adjustments to the dielectric constant can be made by varying temperature and pressure, allowing for optimization opportunities to selectively extract compounds of interest (Gbashi et al., 2017; Mendiola et al., 2013). The high temperature conditions achieved by subcritical water extractions leads to better matrix disruption, and high pressure forces water into the matrix, making extractions more effective (Pillot et al., 2019).

subcritical water extraction methods are gaining popularity as researchers and consumers alike are increasingly interested in the sustainability of food systems. To date, there are many studies utilizing subcritical water extraction to extract mono-, oligo-, and polysaccharide from food byproducts such as grape seeds, coconut husks, fruit peels, and pea hulls (Khuwijitjaru et al., 2014; Klinchongkon et al., 2017; Prado et al., 2014; Ramirez et al., 2021). However, to our knowledge, extraction of oligosaccharides and the release of potentially prebiotic oligosaccharides from enzyme-assisted byproducts of beans and lentils is wholly novel.

Microwave extraction

The utilization of microwave-assisted aqueous extraction has high potential in enhancing the extractability of plant materials by targeting the plant cell wall, which is a key limiting factor in plant processing. Microwave-assisted extractions achieve improved extractability through rapid heating which increases cell matrix degradation and reduces extraction time (Eskilsson and Björklund, 2000). When a polar aqueous solvent is exposed to microwave irradiation, water molecules within the solvent experience vibrations directed by the microwave frequency. These microwaves induce movement and generate friction among the molecules within the cell matrix, resulting in heat production. Consequently, the cell wall matrix is broken down, facilitating increased mass transfer (Bordoloi and Goosen, 2020).

Microwave processing, in combination with conventional solvents, has demonstrated the ability to significantly enhance the extractability of compounds bound to the cell wall matrix, such as polyphenols, with improvements of up to 70% reported (Bordoloi and Goosen, 2020; Magnusson et al., 2017). Additionally, microwave-assisted aqueous extractions have been shown to significantly reduce extraction times from hours to minutes and minimize solvent volume compared to conventional extraction methods (Yuan and Macquarrie, 2015). The advantages offered by microwave-assisted extraction, including enhanced extraction efficiency, reduced processing time, and decreased solvent usage, make this technology a valuable green extraction approach.

Application of green extraction technologies in the presented work

This work features a variety of green extraction technologies. For all plants, aqueous extraction processes were employed, avoiding the use of conventional organic solvents. Proteomic analysis was employed to identify the specific proteins extracted through AEP and EAEP. This research aims to shed light on the protein composition of black beans, lentils and macroalgae across different green extraction methods, facilitating the development of efficient and sustainable approaches for obtaining high-quality bean protein ingredients.

For beans and lentils, subcritical water extraction was employed to generate potentially prebiotic oligosaccharides from insoluble byproducts remaining after enzyme-assisted extraction. Glycomics was utilized to characterize the oligosaccharides released from various conditions of subcritical water extraction.

For macroalgae, microwave-assisted aqueous extraction was employed to extract proteins and carbohydrates. This work explores the potential of microwave-assisted extraction by analyzing

the biomass components (protein, oligosaccharides, monosaccharides) that result from this technology.

References

Agrawal, G. K., Sarkar, A., Righetti, P. G., Pedreschi, R., Carpentier, S., Wang, T., Rakwal, R. (2013). A decade of plant proteomics and mass spectrometry: translation of technical advancements to food security and safety issues. *Mass Spectrometry Reviews*, *32*(5), 335–365. https://doi.org/10.1002/mas.21365

An, H. J., and Lebrilla, C. B. (2010). A glycomics approach to the discovery of potential cancer biomarkers. *Methods in Molecular Biology*, *600*, 199–213. https://doi.org/10.1007/978-1-60761-454-8_14

Baginsky, S. (2009). Plant proteomics: concepts, applications, and novel strategies for data interpretation. *Mass Spectrometry Reviews*, *28*(1), 93–120. https://doi.org/10.1002/mas.20183

Barile, D., and Rastall, R. A. (2013). Human milk and related oligosaccharides as prebiotics. *Current Opinion in Biotechnology*, *24*(2), 214–219. https://doi.org/10.1016/j.copbio.2013.01.008

Bordoloi, A., and Goosen, N. (2020). Green and integrated processing approaches for the recovery of high-value compounds from brown seaweeds. In *Advances in Botanical Research*: *Vol. 95. Seaweeds around the world: state of art and perspectives* (pp. 369–413). Elsevier. https://doi.org/10.1016/bs.abr.2019.11.011

Bose, S. K., Howlader, P., Wang, W., and Yin, H. (2021). Oligosaccharide is a promising natural preservative for improving postharvest preservation of fruit: A review. *Food Chemistry*, *341*(Pt 1), 128178. https://doi.org/10.1016/j.foodchem.2020.128178

Boye, J., Zare, F., and Pletch, A. (2010). Pulse proteins: Processing, characterization, functional properties and applications in food and feed. *Food Research International*, *43*(2), 414–431. https://doi.org/10.1016/j.foodres.2009.09.003

Buckeridge, M. S., Pessoa dos Santos, H., and Tiné, M. A. S. (2000). Mobilisation of storage cell wall polysaccharides in seeds. *Plant Physiology and Biochemistry*, *38*(1–2), 141–156. https://doi.org/10.1016/S0981-9428(00)00162-5

Campbell, K. A., Vaca-Medina, G., Glatz, C. E., and Pontalier, P.-Y. (2016). Parameters affecting enzyme-assisted aqueous extraction of extruded sunflower meal. *Food Chemistry*, *208*, 245–251. https://doi.org/10.1016/j.foodchem.2016.03.098

Cheng, M.-H., Rosentrater, K. A., Sekhon, J., Wang, T., Jung, S., and Johnson, L. A. (2019). Economic Feasibility of Soybean Oil Production by Enzyme-Assisted Aqueous Extraction Processing. *Food and Bioprocess Technology*, *12*(3), 539–550. https://doi.org/10.1007/s11947-018-2228-9

Chen, S., and Harmon, A. C. (2006). Advances in plant proteomics. *Proteomics*, *6*(20), 5504–5516. https://doi.org/10.1002/pmic.200600143

Couture, G., Vo, T.-T. T., Castillo, J. J., Mills, D. A., German, J. B., Maverakis, E., and Lebrilla, C. B. (2021). Glycomic mapping of the maize plant points to greater utilization of the entire plant. *ACS Food Science & Technology*. https://doi.org/10.1021/acsfoodscitech.1c00318

De Angelis, M., Minervini, F., Caputo, L., Cassone, A., Coda, R., Calasso, M. P., … Gobbetti, M. (2008). Proteomic analysis by two-dimensional gel electrophoresis and starch characterization of Triticum turgidum L. var. durum cultivars for pasta making. *Journal of Agricultural and Food Chemistry*, *56*(18), 8619–8628. https://doi.org/10.1021/jf8008876

De Leoz, M. L. A., Kalanetra, K. M., Bokulich, N. A., Strum, J. S., Underwood, M. A., German, J. B., … Lebrilla, C. B. (2015). Human milk glycomics and gut microbial genomics in infant feces show a correlation between human milk oligosaccharides and gut microbiota: a proof-of-concept study. *Journal of Proteome Research*, *14*(1), 491–502. https://doi.org/10.1021/pr500759e

de Moura, J. M. L. N., Campbell, K., Mahfuz, A., Jung, S., Glatz, C. E., and Johnson, L. (2008). Enzyme-Assisted Aqueous Extraction of Oil and Protein from Soybeans and Cream De-emulsification. *Journal of the American Oil Chemists' Society*, *85*(10), 985–995. https://doi.org/10.1007/s11746-008-1282-2

de Moura, Juliana Maria Leite Nobrega, de Almeida, N. M., and Johnson, L. A. (2009). Scale-up of Enzyme-Assisted Aqueous Extraction Processing of Soybeans. *Journal of the American Oil Chemists' Society*, *86*(8), 809–815. https://doi.org/10.1007/s11746-009-1406-3

Ding, W., Nothaft, H., Szymanski, C. M., and Kelly, J. (2009). Identification and quantification of glycoproteins using ion-pairing normal-phase liquid chromatography and mass spectrometry. *Molecular & Cellular Proteomics*, *8*(9), 2170–2185. https://doi.org/10.1074/mcp.M900088-MCP200

Eskilsson, C. S., and Björklund, E. (2000). Analytical-scale microwave-assisted extraction. *Journal of Chromatography. A*, *902*(1), 227–250. https://doi.org/10.1016/s0021-9673(00)00921-3

Gallardo, J. M., Carrera, M., and Ortea, I. (2013). Proteomics in food science. In A. Cifuentes (Ed.), *Foodomics: advanced mass spectrometry in modern food science and nutrition* (pp. 125–165). Hoboken, NJ, USA: John Wiley & Sons, Inc. https://doi.org/10.1002/9781118537282.ch5

Gbashi, S., Adebo, O. A., Piater, L., Madala, N. E., and Njobeh, P. B. (2017). Subcritical water extraction of biological materials. *Separation & Purification Reviews*, *46*(1), 21–34. https://doi.org/10.1080/15422119.2016.1170035

Goli, M., Yu, A., Cho, B. G., Gautam, S., Wang, J., Gutierrez-Reyes, C. D., … Mechref, Y. (2021). LC-MS/MS in glycomics and glycoproteomics analyses. In *Carbohydrate analysis by modern liquid phase separation techniques* (pp. 391–441). Elsevier. https://doi.org/10.1016/B978-0-12-821447-3.00005-6

Haghighi, A., and Khajenoori, M. (2013). Subcritical Water Extraction. In H. Nakajima (Ed.), *Mass Transfer - Advances in Sustainable Energy and Environment Oriented Numerical Modeling*. InTech. https://doi.org/10.5772/54993

Hajduch, M., Agrawal, G. K., and Thelen, J. J. (2007). Proteomics of seed development in oilseed crops. In J. Šamaj and J. J. Thelen (Eds.), *Plant Proteomics* (pp. 137–154). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-72617-3_10

Huang, Y.-P., Robinson, R. C., and Barile, D. (2022). Food glycomics: Dealing with unexpected degradation of oligosaccharides during sample preparation and analysis. *Journal of Food and Drug Analysis*, *30*(1), 62–76. https://doi.org/10.38212/2224-6614.3393

Huang, Y.-P., Robinson, R. C., Dias, F. F. G., de Moura Bell, J. M. L. N., and Barile, D. (2022). Solid-Phase Extraction Approaches for Improving Oligosaccharide and Small Peptide Identification with Liquid Chromatography-High-Resolution Mass Spectrometry: A Case Study on Proteolyzed Almond Extract. *Foods*, *11*(3), 340. https://doi.org/10.3390/foods11030340

Hurkman, W. J., and Tanaka, C. K. (2007). High-Resolution Two-Dimensional Gel Electrophoresis: A Cornerstone of Plant Proteomics. In J. Šamaj and J. J. Thelen (Eds.), *Plant Proteomics* (pp. 14–28). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-72617-3_2

Illippangama, A. U., Jayasena, D. D., Jo, C., and Mudannayake, D. C. (2022). Inulin as a functional ingredient and their applications in meat products. *Carbohydrate Polymers*, *275*, 118706. https://doi.org/10.1016/j.carbpol.2021.118706

Jarpa-Parra, M., Bamdad, F., Wang, Y., Tian, Z., Temelli, F., Han, J., and Chen, L. (2014). Optimization of lentil protein extraction and the influence of process pH on protein structure and functionality. *LWT - Food Science and Technology*, *57*(2), 461–469. https://doi.org/10.1016/j.lwt.2014.02.035

Jorrin-Novo, J. V., Valledor, L., Castillejo, M. A., and Rey, M.-D. (Eds.). (2020). *Plant proteomics: methods and protocols*. New York, NY: Springer US. https://doi.org/10.1007/978-1-0716-0528-8

Joshi, M., Adhikari, B., Aldred, P., Panozzo, J. F., and Kasapis, S. (2011). Physicochemical and functional properties of lentil protein isolates prepared by different drying methods. *Food Chemistry*, *129*(4), 1513–1522. https://doi.org/10.1016/j.foodchem.2011.05.131

Keegstra, K. (2010). Plant cell walls. *Plant Physiology*, *154*(2), 483–486. https://doi.org/10.1104/pp.110.161240

Khuwijitjaru, P., Pokpong, A., Klinchongkon, K., and Adachi, S. (2014). Production of oligosaccharides from coconut meal by subcritical water treatment. *International Journal of Food Science & Technology*, *49*(8), 1946–1952. https://doi.org/10.1111/ijfs.12524

Klinchongkon, K., Khuwijitjaru, P., Wiboonsirikul, J., and Adachi, S. (2017). Extraction of Oligosaccharides from Passion Fruit Peel by Subcritical Water Treatment. *Journal of Food Process Engineering*, *40*(1), e12269. https://doi.org/10.1111/jfpe.12269

Lannoo, N., Van Damme, E. J. M., Albenne, C., and Jamet, E. (2014). Plant Glycobiology-a diverse world of lectins, glycoproteins, glycolipids and glycans. *Frontiers in Plant Science*, *5*, 604. https://doi.org/10.3389/fpls.2014.00604

Lee, H., Htoon, A., and Paterson, J. (2007). Alkaline extraction of starch from Australian lentil cultivars Matilda and Digger optimised for starch yield and starch and protein quality. *Food Chemistry*, *102*(3), 551–559. https://doi.org/10.1016/j.foodchem.2006.03.042

Magnusson, M., Yuen, A. K. L., Zhang, R., Wright, J. T., Taylor, R. B., Maschmeyer, T., and de Nys, R. (2017). A comparative assessment of microwave assisted (MAE) and conventional solid-liquid (SLE) techniques for the extraction of phloroglucinol from brown seaweed. *Algal Research*, *23*, 28–36. https://doi.org/10.1016/j.algal.2017.01.002

Mahajan, R., Nazir, M., Dar, J. A., Mukhtar, S., Mushtaq, M., Sharma, S., … Zargar, S. M. (2021). Glycome profiling in plants: towards understanding the scenario of carbohydrates. In A. Malik and T. A. Dar (Eds.), *The glycome: understanding the diversity and complexity of glycobiology* (pp. 235–273). First edition.: Apple Academic Press. https://doi.org/10.1201/9781003145394-9

Mena-García, A., Ruiz-Matute, A. I., Soria, A. C., and Sanz, M. L. (2019). Green techniques for extraction of bioactive carbohydrates. *TrAC Trends in Analytical Chemistry*, *119*, 115612. https://doi.org/10.1016/j.trac.2019.07.023

Mendiola, J. A., Castro-Puyana, M., Herrero, M., and Ibáñez, E. (2013). Green Foodomics. In A. Cifuentes (Ed.), *Foodomics: advanced mass spectrometry in modern food science and nutrition* (pp. 471–506). Hoboken, NJ, USA: John Wiley & Sons, Inc. https://doi.org/10.1002/9781118537282.ch18

Moreira, A. S. P., Nunes, F. M., Domingues, M. R. M., and Coimbra, M. A. (2015). Galactomannans in Coffee. In *Coffee in health and disease prevention* (pp. 173–182). Elsevier. https://doi.org/10.1016/B978-0-12-409517-5.00019-X

Muszyński, A., O'Neill, M. A., Ramasamy, E., Pattathil, S., Avci, U., Peña, M. J., … Carlson, R. W. (2015). Xyloglucan, galactomannan, glucuronoxylan, and rhamnogalacturonan I do not have identical structures in soybean root and root hair cell walls. *Planta*, *242*(5), 1123–1138. https://doi.org/10.1007/s00425-015-2344-y

Nagai, K., Yotsukura, N., Ikegami, H., Kimura, H., and Morimoto, K. (2008). Protein extraction for 2-DE from the lamina of Ecklonia kurome (laminariales): recalcitrant tissue containing high levels of viscous polysaccharides. *Electrophoresis*, *29*(3), 672–681. https://doi.org/10.1002/elps.200700461

Nandita, E., Bacalzo, N. P., Ranque, C. L., Amicucci, M. J., Galermo, A., and Lebrilla, C. B. (2021). Polysaccharide identification through oligosaccharide fingerprinting. *Carbohydrate Polymers*, *257*, 117570. https://doi.org/10.1016/j.carbpol.2020.117570

Oliveira, D. M., Mota, T. R., Salatta, F. V., Sinzker, R. C., Končitíková, R., Kopečný, D., … Dos Santos, W. D. (2020). Cell wall remodeling under salt stress: Insights into changes in polysaccharides, feruloylation, lignification, and phenolic metabolism in maize. *Plant, Cell & Environment*, *43*(9), 2172–2191. https://doi.org/10.1111/pce.13805

Osborne, T. B. (1924). The Vegetable Proteins. *Nature*, *114*(2875), 822–822. https://doi.org/10.1038/114822c0

Patwa, T. H., Qiu, Y., Zhao, J., Simeone, D. M., and Lubman, D. M. (2009). All-liquid separations, protein microarrays, and mass spectrometry to interrogate serum proteomes: an application to serum glycoproteomics. *Methods in Molecular Biology*, *520*, 75–87. https://doi.org/10.1007/978-1-60327-811-9_6

Pavlovic, M., Huber, I., Konrad, R., and Busch, U. (2013). Application of MALDI-TOF MS for the Identification of Food Borne Bacteria. *The Open Microbiology Journal*, *7*, 135–141. https://doi.org/10.2174/1874285801307010135

Peng, X., and Yao, Y. (2017). Carbohydrates as fat replacers. *Annual Review of Food Science and Technology*, *8*, 331–351. https://doi.org/10.1146/annurev-food-030216-030034

Pillot, M., Lebeau, B., Nouali, H., Daou, T. J., Patarin, J., and Ryzhikov, A. (2019). High pressure intrusion of water and LiCl aqueous solutions in hydrophobic KIT-6 mesoporous silica: Influence of the grafted group nature. *Microporous and Mesoporous Materials*, *280*, 248–255. https://doi.org/10.1016/j.micromeso.2019.02.006

Pischetsrieder, M., and Baeuerlein, R. (2009). Proteome research in food science. *Chemical Society Reviews*, *38*(9), 2600–2608. https://doi.org/10.1039/b817898b

Prado, J. M., Forster-Carneiro, T., Rostagno, M. A., Follegatti-Romero, L. A., Maugeri Filho, F., and Meireles, M. A. A. (2014). Obtaining sugars from coconut husk, defatted grape seed, and pressed palm fiber by hydrolysis with subcritical water. *The Journal of Supercritical Fluids*, *89*, 89–98. https://doi.org/10.1016/j.supflu.2014.02.017

Ramirez, C. S. V., Temelli, F., and Saldaña, M. D. A. (2021). Production of pea hull soluble fiber-derived oligosaccharides using subcritical water with carboxylic acids. *The Journal of Supercritical Fluids*, 105349. https://doi.org/10.1016/j.supflu.2021.105349

Rosenthal, A., Pyle, D. L., and Niranjan, K. (1998). Simultaneous Aqueous Extraction of Oil and Protein from Soybean: Mechanisms for Process Design. *Food and Bioproducts Processing*, *76*(4), 224–230. https://doi.org/10.1205/096030898532124

Šamaj, J., and Thelen, J. J. (Eds.). (2007). *Plant Proteomics*. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-72617-3

Schädel, C., Blöchl, A., Richter, A., and Hoch, G. (2010). Quantification and monosaccharide composition of hemicelluloses from different plant functional types. *Plant Physiology and Biochemistry*, *48*(1), 1–8. https://doi.org/10.1016/j.plaphy.2009.09.008

Serin, S., and Sayar, S. (2016). The effect of the replacement of fat with carbohydrate-based fat replacers on the dough properties and quality of the baked pogaca: a traditional high-fat bakery product. *Food Science and Technology*, *37*(1), 25–32. https://doi.org/10.1590/1678-457x.05516

Singh, D., Chaudhary, P., Taunk, J., Singh, C. K., Singh, D., Tomar, R. S. S., … Pal, M. (2021). Fab advances in fabaceae for abiotic stress resilience: from “omics” to artificial intelligence. *International Journal of Molecular Sciences*, *22*(19). https://doi.org/10.3390/ijms221910535

Souza Almeida, F., Furlan Goncalves Dias, F., Kawazoe Sato, A. C., and Leite Nobrega de Moura Bell, J. M. (2021). From solvent extraction to the concurrent extraction of lipids and

proteins from green coffee: An eco-friendly approach to improve process feasibility. *Food and Bioproducts Processing*, *129*, 144–156. https://doi.org/10.1016/j.fbp.2021.08.004

Sriamornsak, P. (2003). Chemistry of pectin and its pharmaceutical uses: A review. *Silpakorn University International Journal*, *3*, 206–228.

Tang, C.-H., and Sun, X. (2011). Structure–physicochemical function relationships of 7S globulins (vicilins) from red bean (Phaseolus angularis) with different polypeptide constituents. *Food Hydrocolloids*, *25*(3), 536–544. https://doi.org/10.1016/j.foodhyd.2010.08.009

Tang, W., Liu, D., and Nie, S.-P. (2022). Food glycomics in food science: recent advances and future perspectives. *Current Opinion in Food Science*, *46*, 100850. https://doi.org/10.1016/j.cofs.2022.100850

Tan, E.-S., Ying-Yuan, N., and Gan, C.-Y. (2014). A comparative study of physicochemical characteristics and functionalities of pinto bean protein isolate (PBPI) against the soybean protein isolate (SPI) after the extraction optimisation. *Food Chemistry*, *152*, 447–455. https://doi.org/10.1016/j.foodchem.2013.12.008

Toby, T. K., Fornelli, L., Srzentić, K., DeHart, C. J., Levitsky, J., Friedewald, J., and Kelleher, N. L. (2019). A comprehensive pipeline for translational top-down proteomics from a single blood draw. *Nature Protocols*, *14*(1), 119–152. https://doi.org/10.1038/s41596-018-0085-7

Varki, A. (2017). Biological roles of glycans. *Glycobiology*, *27*(1), 3–49. https://doi.org/10.1093/glycob/cww086

Voiniciuc, C., Pauly, M., and Usadel, B. (2018). Monitoring polysaccharide dynamics in the plant cell wall. *Plant Physiology*, *176*(4), 2590–2600. https://doi.org/10.1104/pp.17.01776

Vulevic, J., Drakoularakou, A., Yaqoob, P., Tzortzis, G., and Gibson, G. R. (2008). Modulation of the fecal microflora profile and immune function by a novel trans-galactooligosaccharide mixture (B-GOS) in healthy elderly volunteers. *The American Journal of Clinical Nutrition*, *88*(5), 1438–1446. https://doi.org/10.3945/ajcn.2008.26242

Wilkins, M. R., Sanchez, J. C., Gooley, A. A., Appel, R. D., Humphery-Smith, I., Hochstrasser, D. F., and Williams, K. L. (1996). Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnology & Genetic Engineering Reviews*, *13*, 19–50. https://doi.org/10.1080/02648725.1996.10647923

Yadav, S., Yadav, D. K., Yadav, N., and Khurana, S. M. P. (2015). Plant glycomics: advances and applications. In D. Barh, M. S. Khan, and E. Davies (Eds.), *Plantomics: the omics of plant science* (pp. 299–329). New Delhi: Springer India. https://doi.org/10.1007/978-81-322-2172-2_10

Yang, J. S., Dias, F. F. G., Pham, T. T. K., Barile, D., and de Moura Bell, J. M. L. N. (2023). A sequential fractionation approach to understanding the physicochemical and functional properties of aqueous and enzyme-assisted aqueous extracted black bean proteins. Submitted for publication to *Food Hydrocolloids*.

Yates, J. R. (2004). Mass spectral analysis in proteomics. *Annual Review of Biophysics and*

Biomolecular Structure, *33*, 297–316. https://doi.org/10.1146/annurev.biophys.33.111502.082538

Yuan, Y., and Macquarrie, D. (2015). Microwave assisted extraction of sulfated polysaccharides (fucoidan) from Ascophyllum nodosum and its antioxidant activity. *Carbohydrate Polymers*, *129*, 101–107. https://doi.org/10.1016/j.carbpol.2015.04.057

Zadražnik, T., Moen, A., Egge-Jacobsen, W., Meglič, V., and Šuštar-Vozlič, J. (2017). Towards a better understanding of protein changes in common bean under drought: A case study of N-glycoproteins. *Plant Physiology and Biochemistry*, *118*, 400–412. https://doi.org/10.1016/j.plaphy.2017.07.004

Zhang, J., Wen, C., Zhang, H., Duan, Y., and Ma, H. (2020). Recent advances in the extraction of bioactive compounds with subcritical water: A review. *Trends in Food Science & Technology*, *95*, 183–195. https://doi.org/10.1016/j.tifs.2019.11.018

Zhang, Q. (2017, September). *ETX220/FOR220 – Analysis of Toxicants*. Presented at the Lecture, University of California, Davis.

Zhang, Y., Fonslow, B. R., Shan, B., Baek, M.-C., and Yates, J. R. (2013). Protein analysis by shotgun/bottom-up proteomics. *Chemical Reviews*, *113*(4), 2343–2394. https://doi.org/10.1021/cr3003533

Zhao, X., Zeng, L., Wang, J., Shi, Y., Zhang, B., Liu, Y., … Li, X. (2023). Quantitative N-Glycomic and N-Glycoproteomic Profiling of Peach [Prunus persica (L.) Batsch] during Fruit Ripening. *Journal of Proteome Research*, *22*(3), 885–895. https://doi.org/10.1021/acs.jproteome.2c00662

Zivkovic, A. M., and Barile, D. (2011). Bovine milk as a source of functional oligosaccharides for improving human health. *Advances in Nutrition (Bethesda, Md.)*, *2*(3), 284–289. https://doi.org/10.3945/an.111.000455

CHAPTER 2:

Elucidatating black bean (*Phaseolus vulgaris*) protein and carbohydrate composition and the effects of green extraction methods

A modified version of this chapter has been submitted for publication to Food Hydrocolloids.[1]

Abstract

Despite the nutritional value and sustainable advantages of dry common beans (*Phaseolus vulgaris*), their potential as a high-value ingredient in food products remains underutilized, particularly in comparison to other legumes. This study aims to address this research gap by investigating the influence of green extraction methods on the protein and carbohydrate composition of black beans using proteomics and glycomics methods. Proteomics analysis explored various protein compositions resulting from different extraction techniques, including aqueous and enzyme-assisted aqueous extractions. Enzyme-assisted extraction significantly improved protein yields, and proteomics results showed differences between aqueous extraction and enzyme-assisted aqueous extraction, confirming the impact of proteolysis on the protein composition. Further, glycomics techniques characterized the oligosaccharides released through subcritical water extraction of insoluble bean byproduct remaining after enzyme-assisted extraction. Extraction temperature and time significantly influenced the number and composition of identified oligosaccharides. Higher temperatures yielded a greater number of oligosaccharides, while lower temperatures favored the release of specific hexose-only oligosaccharides. Extended extraction times at lower temperatures were required to extract the same oligosaccharides, but pre-equilibrium fractions consistently exhibited the most abundant and diverse glycoprofiles. The library generated for glycomics analysis offers insights into oligosaccharide release during subcritical water extraction, providing opportunities for process optimization. This research enhances the understanding of black bean proteins and carbohydrates, shedding light on the effects of different extraction methods. Proteomics and

[1] Yang, J. S., Dias, F. F. G., Pham, T. T. K., Barile, D., and de Moura Bell, J. M. L. N. (2023). A sequential fractionation approach to understanding the physicochemical and functional properties of aqueous and enzyme-assisted aqueous extracted black bean proteins. Submitted for publication to *Food Hydrocolloids*.

glycomics techniques offer valuable insights into the impacts of green extraction techniques, such as enzyme-assisted extraction and subcritical water extraction, on black bean components. These findings facilitate the development of innovative food products and contribute to sustainable food production practices, promoting the utilization of black beans as a high-value ingredient.

Introduction

The dry common bean (*Phaseolus vulgaris*) has long been integral to diets worldwide, serving as a crucial staple food and an essential source of nutrition for the growing global population. Beans are legumes, which account for 27% of global primary crop production (Duc et al., 2015). Legumes are categorized into two classes: oilseeds (e.g. soybeans and peanuts) and grain legumes (e.g. common beans, lentils, peas) (Geil and Anderson, 1994). Dried grain legumes, like beans and lentils, are also known as “pulses” (USA Pulses, 2018). Recognizing their value, the United Nations General Assembly declared 2016 as the International Year of Pulses (United Nations General Assembly, 2014). Pulses are valuable not only for their nutritional content, but also for their agricultural and environmental benefits. Due to their nitrogen-fixing abilities, cultivating pulses benefits soil health, increases farmland productivity, and reduces reliance on fertilizers which release greenhouse gasses during manufacturing and application (United Nations, 2023). Furthermore, pulses are water efficient, requiring less than 10% of the water necessary for beef production (FAO, 2015).

Dry beans are high in carbohydrates (starch, dietary fibers), proteins, and minerals, and they are low in fat (de Almeida Costa et al., 2006; Kutoš et al., 2003). This unique composition allows dry beans to be categorized into two of the six core elements making up a healthy diet, as established in US Dietary Guidelines for Americans (2020-2025) (U.S. Department of Agriculture and U.S. Department of Health and Human Services, 2020). With their high protein

content, dry beans are categorized as a valuable non-animal derived source of protein and simultaneously, they are categorized as a nutrient-dense vegetable due to their high fiber content.

Legumes and pulses in general are an excellent source of plant proteins, containing 2-3 times that of cereal grains (Siddiq et al., 2010). Despite their favorable composition, utilization of bean proteins in plant-based meat alternatives, snacks, and bakery products lags behind that of other legume (soy) and pulse (pea) proteins.

Bean composition

Beans are primarily made of carbohydrates which make up about 68% of the dry matter, and protein which make up about 25% of the dry matter (Berrios et al., 1999). Proximate analysis of black bean flour utilized in this work showed 11.3 ± 0.3% moisture, 20.0 ± 0.1% protein, 2.3 ± 0.1% oil, 3.8 ± 0.4% ash, and 62.6 ± 0.2% total carbohydrates (J. S. Yang et al., 2023).

Bean proteins

Black bean's relatively high protein content makes it highly valuable in global human consumption, and their production helps to address the issue of protein malnutrition. Proteins constitute the second most abundant component of beans, accounting for approximately 25% of their dry weight (Berrios et al., 1999). This protein content is comparable to that of meat (Almeida et al., 2006). While oilseed proteins, notably soybeans, have been extensively studied and utilized in food products, the characterization of bean proteins has been lacking (Sathe, 2002).

Similarly to other legumes, bean proteins are rich in essential amino acids such as lysine, phenylalanine, and tyrosine, but they are lower in sulfur-containing amino acids, methionine and

cysteine (Sathe, 2002). This nutritional profile complements the deficiency of cereal grain proteins, making beans and cereal proteins nutritionally complementary. The consumption of a combination of beans and cereals ensures a complete protein with all essential amino acids.

Bean proteins can be categorized based on function or structure. Based on function, bean proteins can be classified into “storage, carbohydrate metabolism, defense, stress response, detoxification, growth and development, protein transport and nitrogen metabolism” (Luna-Vital et al., 2015). Based on structure, bean proteins can be fractionated into globulins, albumins, prolamin, and glutelin (Hayat et al., 2014). Globulins are storage proteins making up 50-70% of total bean proteins; these are the largest protein fraction in beans (Sathe, 2002). Albumins, water soluble proteins, represent the second most abundant class of bean proteins and make up 10-30% of total protein content by dry weight.

Globulins are classified into two types, 7S and 11S, which are differentiated by the number of subunits comprising the protein. The 7S globulins in beans, commonly referred to as phaseolin, are trimeric proteins made up of three subunits, and the 11S globulins are hexameric proteins made up of six subunits (Kimura et al., 2008). The subunits making up 7S or 11S globulins vary by amino acid sequence, molecular weight, degree of glycosylation, and isoelectric points, resulting in diverse physicochemical properties (Tang and Sun, 2011).

Unprocessed bean proteins are notoriously resistant to proteolysis and digestion in humans (Hayat et al., 2014; Nielsen et al., 1988). The abundance of β-sheets (a common type of secondary structural motif in proteins) in the structures of 7S and 11S proteins causes these proteins to be highly rigid and compact, limiting the access of digestive proteolytic enzymes and leading to decreased digestibility in the human gastrointestinal tract (Deshpande and Damodaran, 1989; Yu, 2005). Additionally, protein conjugates such as glycoproteins, contribute

to increased proteolysis resistance (Genovese and Lajolo, 1996). Thermal treatment of bean proteins induces structural changes that may increase the digestibility and biological value of these proteins (Nergiz and Gökgöz, 2007). Understanding the effects of various extraction methods, the resulting protein profiles, and the degree of protein hydrolysis achieved by these extractions is essential for the utilization of bean proteins in food products.

Bean carbohydrates

Bean carbohydrates consist largely of complex structural and storage polysaccharides (Geil and Anderson, 1994), including starch and nonstarch polysaccharides. Starch, the major carbohydrate, exists in two major forms: amylose and amylopectin. Amylose is a linear polysaccharide composed of glucose units with α-glycosidic bonds, coiled into a spiral structure. Amylopectin, on the other hand, is highly branched and consists of 20-25 glucose units (Green et al., 1975). Generally, beans contain a higher proportion of amylopectin than amylose (Reddy et al., 1984).

Nonstarch polysaccharides, also known as dietary fiber, can be classified into soluble and insoluble fiber. Soluble fibers include pectins and oligosaccharides, which contribute to lowering blood glucose and blood glucose regulation (Guillon and Champ, 2002; Slavin, 2013). Insoluble fiber, including cellulose, hemicellulose, and lignin, resists digestion in the small intestine and undergoes fermentation in the large intestine (Dhingra et al., 2012). Beans contain a high proportion of non-digestible carbohydrates, including non-digestible oligosaccharides, resistant starch, and soluble and insoluble dietary fiber (Henningsson, Nyman, and Björck, 2001). Several types of prebiotics, such as resistant starch and fructooligosaccharides, have been identified in beans (Câmara et al., 2013).

Beans also contain smaller amounts of mono-, di-, and oligosaccharides (Bravo et al. 1998). Raffinose family oligosaccharides, including raffinose, stachyose, and verbascose, are present in varying amounts in beans (Geil and Anderson, 1994). These oligosaccharides require the enzyme alpha-galactosidase for hydrolysis, which is absent in the human digestive system. Consequently, raffinose family oligosaccharides remain undigested in the small intestine and undergo microbial fermentation in the large intestine, resulting in gas production and flatulence (Elango et al., 2022). To mitigate these effects, raffinose family oligosaccharides are commonly removed during bean soaking and cooking processes for at home consumption and industrial processing.

Despite the somewhat undesirable gas production in the human intestine, consumption of beans can also elicit beneficial physiological responses, such as modulation of gastrointestinal transit time, reduction of cholesterol levels, increased satiety and glycemic control (Anderson et al., 1990; de Almeida Costa et al., 2006; Guillon and Champ, 2002; Park et al., 2004). For these reasons, resistant starch and dietary fiber play a crucial role in managing metabolic syndrome, which is associated with an elevated risk of cardiovascular disease (Hayat et al., 2014). These benefits are attributed to the production of short-chain fatty acids (SCFA) that result from microbial fermentation (Finley et al. 2007). Dietary fibers, in general, help to slow the release of carbohydrates during digestion, and their ingestion has been shown to play a role in disease development and management (Guillon and Champ, 2002).

Protein and carbohydrate analysis in the presented work with black beans

One of the major obstacles to incorporating bean proteins as a protein ingredient lies in the inadequate characterization of bean extracts and protein isolates. The work seeks to address this limitation by investigating the potential of black bean as a high-value ingredient, thereby expanding the range of applications for bean proteins. This work employs proteomics

techniques to characterize the changes in protein profiles that result from various extraction conditions, focusing on green extraction methods. In particular, aqueous extractions are explored, and enzyme-assisted extractions are employed with the aim of increasing bean protein extractability and digestibility. Osborne fractionation is used to investigate how protein compositions are influenced by various extraction media, and this method of fractionation serves as a point of comparison in the development of aqueous and enzyme-assisted extractions.

Glycomics techniques are utilized to investigate the breakdown of insoluble byproduct remaining after green extractions of proteins. Following extraction of black bean flour using an enzyme-assisted extraction process, an insoluble fraction is obtained. After removing starch from this insoluble fraction, an uncharacterized byproduct remains. Subcritical water extraction was employed to break down this byproduct to generate potentially prebiotic oligosaccharides from what would typically be considered waste, and glycomics was used to determine the composition of these oligosaccharides.

Experimental procedure

Figure 2.1 provides an overview of the experimental procedures involved in the analysis of bean products, including extraction and analysis methods. Methods relevant to the protein analysis portion of this work are in the top half of the figure, and methods relevant to the carbohydrate analysis of this work are in the bottom half of the figure.

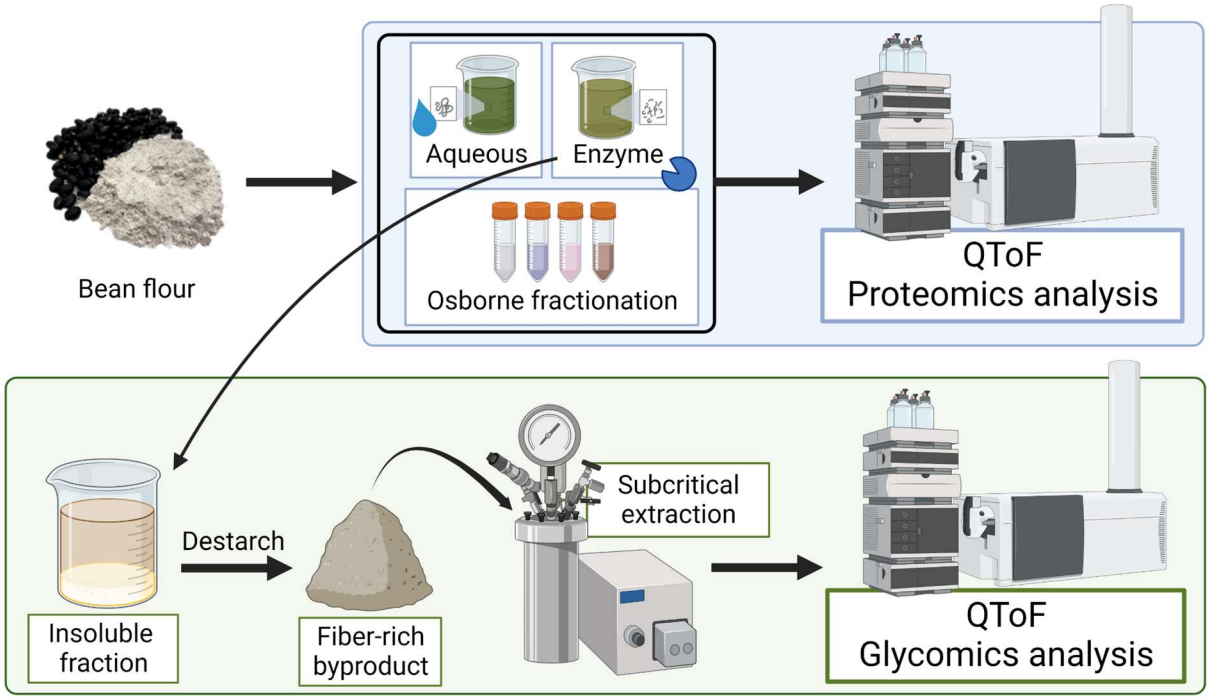


Figure 2.1. Overview of black bean protein (top) and carbohydrate (bottom) extraction and analysis. Created with BioRender.com.

Protein extraction methods

Prior to proteomics analysis, black bean flour was extracted using aqueous extraction (AEP), enzyme-assisted aqueous extraction (EAEP), and Osborne fractionation. These methods are as described by Yang et al. (2023). Briefly, the AEP was performed by mixing 50 g of black bean flour in 500 g of DI water and adjusting the slurry pH to 9.0 with 1 M NaOH. Extraction took place for 1 h under constant stirring (50 °C, 120 rpm) and the pH of the slurry was maintained at pH 9.0. Following extraction, the slurry was centrifuged (4000 x g, 30 min, 4 °C) to separate the protein-rich extract from the starch/fiber-rich insoluble fraction. The EAEP was performed under identical extraction conditions as the AEP, but with the addition of 0.5% (w/w; weight of enzyme/weight of black bean flour) of FoodPro® Alkaline Protease (Danisco, Rochester, NY, USA). Extractions were performed in triplicate and total protein extraction (TPE) were determined using Equation 1.

For Osborne fractionation, black bean flour was sequentially extracted with DI water, 1 M NaCl, 70% ethanol, and 0.05 M NaOH at 25 °C for 1 h in each solvent at a 1:10 solids-to-liquid ratio with constant mixing at 120 rpm to extract albumin, globulin, prolamin, and glutelin, respectively. Following each extraction stage, the slurry was centrifuged (4000 x g, 10 min, 4 °C), and the supernatant (protein-rich extract) was collected. Extractions were performed in triplicate and total protein extraction (TPE) were determined using Equation 1.

Equation 1

$$TPE\ (\%) = [1 - (\frac{Protein\ (g)\ in\ the\ Osborne\ fraction\ or\ final\ insoluble\ fraction)}{Protein\ (g)\ in\ black\ bean\ flour}) \times 100$$

The insoluble fraction (precipitate) was resuspended in water at a 1:10 solids to liquid ratio and extracted for 2 h at 90 °C with 1% w/w FoodPro® AHT, a commercial amylase, to generate a destarched insoluble fraction for subsequent subcritical water experiments.

SDS-PAGE

To visualize the degree of protein hydrolysis and provide context to proteomics analysis, sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) under reducing conditions was performed. These methods are as described by Yang et al. (2023). Briefly, protein samples were diluted 1:1 with 2X Laemmli (5% β-mercaptoethanol) sample buffer (Bio-Rad, Hercules, CA, USA), and samples were heated at 80 °C for 10 min and cooled to RT before loading 35 µg of protein per well to precast protein gels (Any KD™ Criterion™ TGX™ Precast Midi Protein Gel, Bio-Rad, Hercules, CA, USA). Electrophoresis was performed at 200 V. Gels were stained with Bio-Safe™ Coomassie Blue (Bio-Rad, Herculues, CA, USA) for 1 h and destained in DI water. Gels were imaged using a Bio-Rad Gel Doc™ EZ Imager and relative band quantification was performed using Image Lab (Bio-Rad, Hercules, CA, USA).

Proteomics of black bean extracts

Protein precipitation

Extraction replicates were pooled for proteomics analysis. Aside from the prolamin fraction samples which was already extracted under organic conditions, all samples were subjected to protein precipitation as follows:

The pooled protein-extracted samples (400 µL) were mixed with ice-cold trichloroacetic acid (MilliporeSigma, St. Louis, MO)/acetone (Fisher Scientific, Waltham, MA) (10/90 v/v) (1.6 mL) containing 20 mM dithiothreitol (Fisher Scientific, Waltham, MA) added immediately before use. Samples were incubated at -20 °C for 1.5 h, then centrifuged (4,700 rpm, 4 °C, 30 min) (Allegra X-30R Centrifuge, Beckman Coulter, Brea, CA). The supernatant was carefully discarded without disturbing the pellet. The pellet was washed with ice-cold acetone (500 µL), then centrifuged as before. Acetone washes were repeated twice more in the same manner. Following the final removal of supernatant, pellets were left to dry completely of acetone.

Reduction, alkylation, and digestion

Pelleted protein samples were dissolved in 50 mM ammonium bicarbonate (Spectrum Chemical, Gardena, CA) (250 µL). For prolamin fraction samples, ammonium bicarbonate was added to the sample for a final concentration of 50 mM. The equivalent of 50 µg protein (determined by Qubit 3.0 fluorometer assay, ThermoFisher Scientific, Waltham, MA) was transferred to fresh tubes. Concentrated samples were diluted to 1 µg/µL protein using additional 50 mM ammonium bicarbonate. Dithiothreitol was added to a final concentration of 5 mM, and samples were incubated (60 °C, 30 min). Iodoacetamide (MilliporeSigma, St. Louis, MO) was added to a final concentration of 20 mM, and samples were incubated in the dark (room temperature, 30 min). Trypsin (Promega, Madison, WI) was added (1:50 ratio), and samples were incubated (37 °C, overnight) (Eppendorf ThermoMixer C, Enfield, CT). The

following day, trypsin was inactivated by reducing pH to 2-3 using 1% TFA. All samples were centrifuged to pellet insoluble material (14,000 xg, 4 °C, 15 min) (Eppendorf Centrifuge 5424, Enfield, CT).

C18 Cleanup

Digested samples were further purified by microplate C18 solid-phase extraction (SPE) (Glygen, Columbia, MD). Acetonitrile (LC-MS grade, Fisher Scientific, Waltham, MA) and trifluoroacetic acid (TFA) (MilliporeSigma, St. Louis, MO) were used for C18 SPE. The microplate wells were activated with 99.9% acetonitrile/0.1% TFA (v/v) and equilibrated with 1% acetonitrile /0.1% TFA in water (v/v/v). Samples were loaded, and wells were washed with 1% acetonitrile /0.1% TFA in water (v/v/v) (1.2 mL). Peptides were eluted with 80% acetonitrile /0.1% TFA in water (v/v/v) (600 µL). The eluent was dried (Eppendorf Vacufuge plus, Enfield, CT) and redissolved in 3% acetonitrile /0.1% FA in water (v/v/v) for LC-MS/MS analysis.

LC-MS/MS analysis

LC-MS/MS analysis was performed on an Agilent 6520 Accurate-Mass Q-TOF LC-MS with a Chip Cube interface (Agilent Technologies, Santa Clara, CA). Mobile phases contained MilliQ water, acetonitrile (LC-MS grade, Fisher Scientific, Waltham, MA) and formic acid (LC-MS grade, Fisher Scientific, Waltham, MA). Samples were injected onto a Zorbax 300SB-C18, 5 µm, 150 mm x 75 µm Chip with 0.1% FA in water (v/v) at a flow rate of 4 µL min−1. Chromatographic separation of peptides was performed with a gradient consisting of 0.1% FA in water (v/v; A) and 90% acetonitrile, 0.1% FA in water (v/v/v, B) at a flow rate of 0.3 µL min−1. The 80-min gradient was ramped from 0-30% B, 5-60 min; 30-100% B, 60-65 min; 100% B, 65-70 min; 100-0% B, 70 min; 0% B, 70-80 min. The capillary voltage was set to 1950 V. The drying gas was set to 325 °C with a flow rate of 5 L min−1. Mass-to-charge ratio (m/z) was scanned at a rate of 8 spectra sec^{-1} in the m/z range of 275-1700. The precursors were selected

based on abundance and isolated with a width of 1.3 m/z for fragmentation. A ramped collision energy with the equation (0.03 × m/z + 2) was applied to ions of any charges. The MS/MS analysis was scanned at a rate of 0.63 spectra sec^{-1} in the m/z range of 50–1750.

Data analysis for protein identification

PEAKS Studio X+ (Bioinformatics Solutions Inc., Waterloo, ON, Canada) was used for analyzing LC-MS/MS data for peptide and protein identification. Peptides were identified through database search using the Uniprot database with (https://www.uniprot.org/, accessed 1/17/2022), both Swiss-Prot and TrEMBL, with the organism name *Phaseolus vulgaris*. The mass error tolerance was as low as 20 ppm and 0.035 Da for the precursor and fragment ions, respectively. The enzyme was set to “Trypsin” with a specific digestion mode. The number of maximum missed cleavages per peptide was set to 2. Carbamidomethylation was set as a fixed PTM. A maximum of 5 variable modifications, including oxidation (M), phosphorylation (STY), and deamidation (NQ), was allowed. The results were filtered with a false discovery rate of 1.0%. Only proteins with at least 1 unique peptide were retained. Database matches were manually inspected to select the correct protein match.

Glycoprofiling of subcritical extracts

Prior to subcritical water extraction and glycoprofiling, the insoluble fraction remaining after protein extraction underwent destarching by commercial amylase, resulting in an 80% reduction in starch content. To assess the effects of temperature on glycoprofiles, subcritical water extraction was carried out at three temperatures (120 °C, 160 °C, and 200 °C). To assess the effects of extraction time over a 60 min extraction, fractions were collected for pre-equilibrium (the time after the vessel has reached the set temperature but before the target pressure is reached), 0-20 min, 20-40 min, and 40-60 min.

Oligosaccharides released from subcritical extraction were isolated by removing proteins via cold ethanol precipitation. Two volumes of cold ethanol were added to reconstituted fractions and stored at -20°C for 1 h. Samples were centrifuged (4200 xg, 4 °C, 30 min) (Allegra X-30R Centrifuge, Beckman Coulter, Brea, CA). Supernatants were dried by centrifugal evaporation (Genevac miVac, Ipswich, United Kingdom) then reconstituted on water for two stages of solid-phase extraction (SPE). For carbohydrate-normalized analysis, sample volumes were normalized to total soluble carbohydrate content as determined by phenol-sulfuric acid method (DuBois, Gilles, Hamilton, Rebers, and Smith, 1956) to give 20 mg/mL soluble carbohydrate concentration. Samples were cleaned by C18 microplate (Glygen, Columbia, MD), followed by microplate porous graphitized carbon (PGC) SPE (Thermo Fisher Scientific, Waltham, MA). Prior to sample loading onto SPE microplates, reconstituted samples were centrifuged at 14,000 xg, 4°C, for 30 minutes (Eppendorf Centrifuge 5424, Enfield, CT) to remove particulates.

Supernatants were cleaned by SPE as follows. Between all additions, microplates were centrifuged at 1300 rpm, 20°C, for 1 min. C18 microplates were conditioned with 200 µL acetonitrile (x 3), and equilibrated with 200 µL water (x 3). Samples (200 µL) were loaded atop of a fresh collection plate. The flowthrough and subsequent washes (3 x of 200 µL water) were collected for PGC SPE.

PGC microplates were conditioned with 200 µL 80/20 acetonitrile/water containing 0.1% TFA (x 5). Wells were equilibrated with 200 µL water (x 4). The collected flowthrough and washes from C18 SPE were loaded in 200 µL aliquots, centrifuging between additions. Wells were washed with 200 µL water (x 6). A fresh collection plate was used to collect the eluent (3 x 200 µL 40/60 acetonitrile/water containing 0.1% TFA). Eluents were transferred to Eppendorf tubes and dried by centrifugal evaporation. Samples were reconstituted in MilliQ water for LC-MS analysis.

Purified oligosaccharides were analyzed by LC-MS/MS using an Agilent 6520 Accurate-Mass Q-TOF LC-MS with a Chip Cube interface coupled to an Agilent 1200 Series high performance liquid chromatography system (Agilent Technologies, Santa Clara, CA, USA). Mobile phases contained MilliQ water, acetonitrile (LC-MS grade, Fisher Scientific, Waltham, MA) and formic acid (LC-MS grade, Fisher Scientific, Waltham, MA). Specifically, the mobile phases were composed of 0.1 % formic acid (FA), 3% acetonitrile in water (mobile phase A) and 0.1% FA, 90% acetonitrile in water (mobile phase B). Samples containing oligosaccharides were injected onto an Agilent PGC-Chip II (G4240-64010) at a flow rate of 4 µL/min. Oligosaccharides were chromatographically separated over a 60 min gradient at a flow rate of 0.3 µL/min, beginning with 100% A for 2.5 min, ramping from 0% to 16% B in 20 min, increasing from 16% to 44% B in 10 min, then 44% to 100% B in 5 min, and held at 100% B for 10 min. The system was re-equilibrated for 15 min at 100% A prior to the next injection. The drying gas was set at 350 °C with a flow rate of 5 L/min. The electrospray ion source was in positive ion mode with a capillary voltage of 1850 V. The ions were scanned within the range of m/z 150–2500 at a rate of 1 spectrum/sec. The four most abundant ions in each MS analysis cycle were isolated for tandem MS analysis with ramped collision energy (CE; CE = 0.02 × m/z – 3.5). Reference ions m/z 922.009798 and m/z 1221.990637 were used for continual mass calibration throughout the analysis.

Fragmentation data was annotated by Glyconote (https://github.com/MingqiLiu/GlycoNote) and manually inspected using Agilent Masshunter Qualitative Analysis (B.07.00, Agilent Technologies) to identify oligosaccharides structures. Verified oligosaccharides were compiled into a novel library. Peaks were manually integrated by Agilent Masshunter Profinder (B.08.00, Agilent Technologies) using targeted feature extraction which included the monoisotopic masses (with a mass error within 20 ppm) and retention times for all identified oligosaccharides. In-source fragment ions and dimer and trimer aggregates were manually searched and summed

to approximate their actual abundance (Huang et al., 2022). GraphPad Prism (ver. 9.4.0) was used to generate heatmaps from peak area data.

Results

Total protein extracted

The AEP and EAEP achieved total protein extraction yields of 75.0 ± 0.8% and 81.3 ± 0.3%, respectively (Fig. 2.2). The increased total protein extracted can be attributed to the addition of Alkaline Protease. For Osborne fractions, the albumin-rich fraction exhibited the highest extraction of black bean flour protein at 56.2 ± 1.3%, followed by the globulin-rich fraction at 21.9 ± 1.3%, and the glutelin-rich fraction at 15.7 ± 0.6%. The prolamin-rich fraction accounted for less than 1% of the black bean flour protein 0.65 ± 0.03%) (Fig. 2.2). These findings highlight that water alone, under specific conditions (25 °C, 1 h, 1:10 solids-to-liquid ratio), can extract more than half of the total protein present in black bean flour. In total, sequential fractionation facilitated the extraction of approximately 94% of the total protein content.

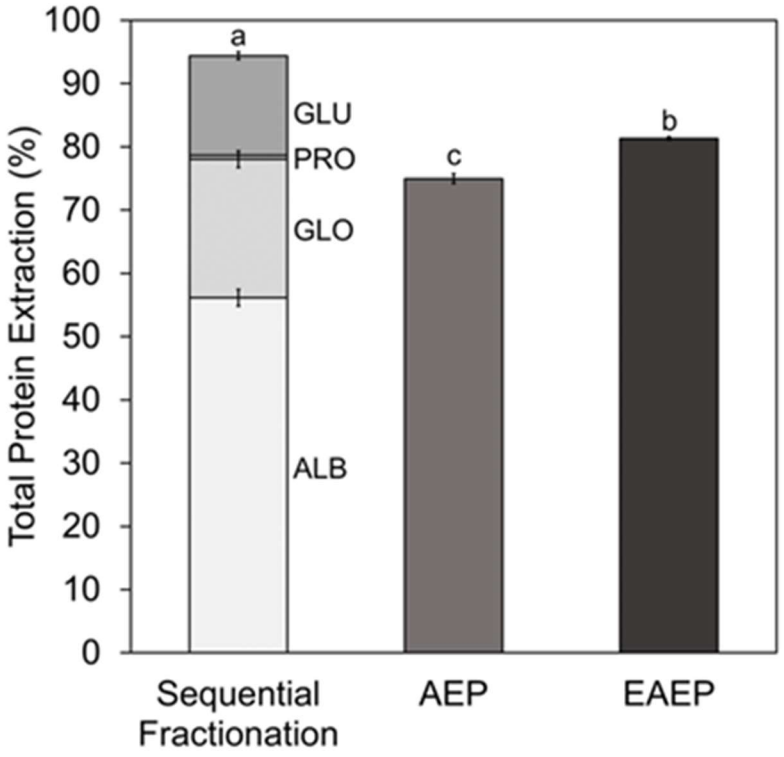


Figure 2.2. Total protein extraction (TPE, %) of the sequential fractionation, aqueous extraction process (AEP), and enzyme-assisted aqueous extraction process (EAEP). The protein distribution of the sequential fractionation was separated into the ALB = albumin-rich, GLO = globulin-rich, PRO = prolamin-rich, and GLU = glutelin-rich fractions. Error bars denote ± SD (n = 3) of the TPE and protein distribution, and different letters indicate statistically significant differences in the TPE by one-way ANOVA followed by Tukey's test ($p < 0.05$).

SDS-PAGE

SDS-PAGE analysis under reducing conditions was employed to assess the impact of extraction media and proteolysis on the molecular weight profiles of extracted proteins (Fig. 2.3). The observed band patterns suggests that the two most prominent bands (~42 and ~48 kDa) in the albumin and globulin lanes (denoted by (*) in Fig. 2.3) could correspond to various subunits of phaseolin, the primary storage protein in beans (Rui et al., 2011). Additionally, the ~31 kDa protein (denoted by (^) in Fig. 2.3) in the albumin and globulin fractions, likely corresponded to PHA (phytohemaglutinnin) (Rui et al., 2011). Notably, phaseolin was more readily extracted in the saline media (globulin), while PHA was primarily extracted during the initial aqueous extraction (albumin), as indicated by the relative band thicknesses.

The SDS-PAGE bands of the AEP proteins represented a combination of the bands observed for the individual protein classes. The significant reduction of the main ~48.5 kDa band (likely α-phaseolin; denoted by (*) in Fig. 2.3) and the emergence of the unique ~26.5 kDa band (denoted by (**) in Fig. 2.3) in the EAEP lane indicated that a significant portion of this protein (~80% as determined by relative band quantification) underwent hydrolysis by alkaline protease, resulting in peptides approximately half its original size. This aligns with previous studies highlighting the susceptibility of the central region of phaseolin to enzymatic hydrolysis (Shpande and Nielsen, 1987; Zhang and Romero, 2020). Furthermore, the loss of the ~31 kDa PHA band in the EAEP lane (denoted by (^) in Fig. 2.3) was consistent with the proteolysis-induced degradation reported for alcalase-hydrolyzed jamapa beans (*P. vulgaris*) (Torruco-Uco et al., 2009). Overall, the addition of alkaline protease during the extraction was highly effective

in proteolysis, as evidenced by the prevalence of proteins/peptides with molecular weights below 26.6 kDa, accounting for 62% of the EAEP extract compared to 13% in the AEP extract.

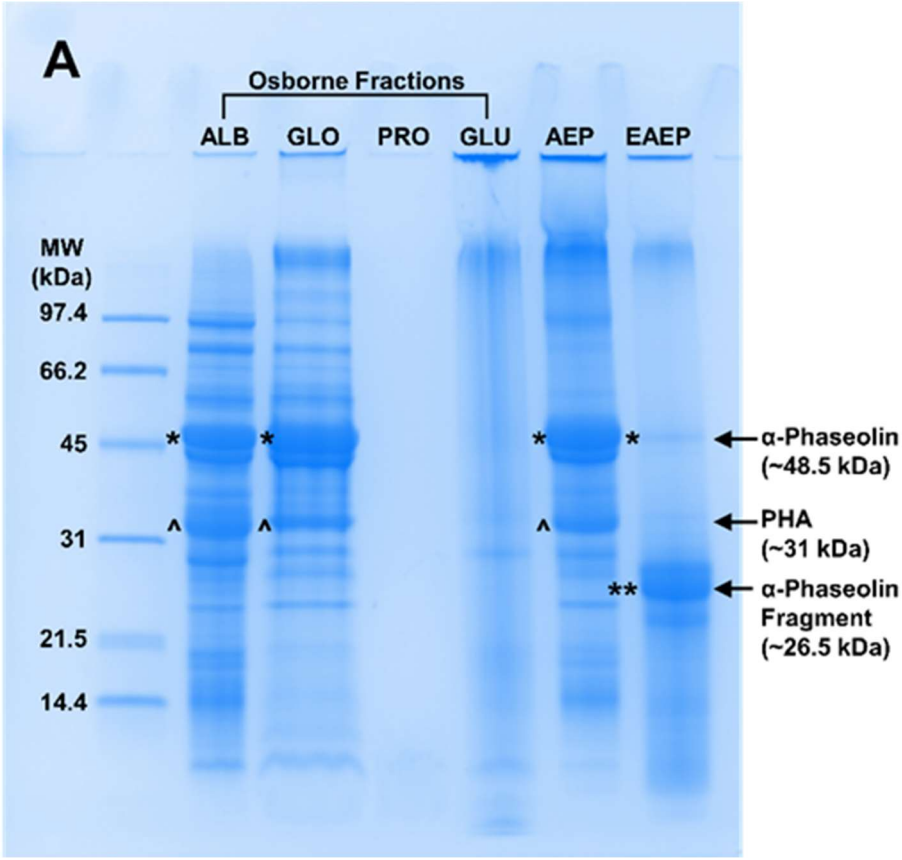


Fig. 2.3. Protein SDS-PAGE gels of the protein classes (ALB = albumin-rich fraction, GLO = globulin-rich fraction, PRO = prolamin-rich fraction, GLU = glutelin-rich fraction), AEP (aqueous extraction process) protein extract, and EAEP (enzyme-assisted aqueous extraction process) protein extract under reducing conditions. The (*) denotes α-phaseolin, while (**) denotes the unique band in the EAEP lane that could be a fragment of α-phaseolin. The (^) denotes PHA (phytohemaglutinnin).

Proteomics results of AEP, EAEP, and Osborne fractions

Liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS) was employed for protein identification to investigate the impact of different extraction media and proteolysis (EAEP) on protein composition. Table 2.1 presents the top 10% most abundant identified proteins, based on peak areas, in each sample. A total of 180 unique proteins were identified using database matching (Supplementary Material Table S2.1). A large proportion of the

reported proteins were identified as hypothetical proteins. These are proteins which are predicted to be expressed based on nucleic acid sequence, but there is not yet experimental evidence supporting their translation or function (Ijaq et al., 2019; Lubec et al., 2005). For the 10% most abundant identified proteins reported in Table 2.1, the proteins were categorized by function (e.g., storage, stress, metabolism) as described by Deb-Choudhury et al. (2021). Proteomic analysis has been previously performed for the entire common bean seed (De La Fuente et al., 2011; Marsolais et al., 2010), but to our knowledge, has not been applied to sequentially extracted protein fractions or enzymatically extracted proteins (EAEP) as accomplished in this work.

Proteomic analysis revealed that the albumin, globulin, and glutelin fractions exhibited similar protein profiles, while the prolamin fraction showed distinct differences (Table 2.1). Despite comparable protein profiles, the relative abundance of the protein species in the fractions varied. For example, arcelin-4 and arcelin 4-II were in the top 10% of proteins in the albumin fraction, but not in the globulin fraction (although still detected). Notably, major proteins in beans including α-phaseolin, phytohemagglutinin, and erythroagglutinating phytohemagglutinin were detected in all of the protein fractions. This is not surprising due to the sequential nature of the fractionation method without further purification. In a previous study that identified proteins in sequentially extracted chickpea albumins, globulins, and glutelins, several proteins (e.g., lipoxygenase, legumin subunits) were also detected in multiple protein fractions (Chang et al. 2011).

The most prevalent protein in all fractions (except for prolamin) was α-phaseolin (MW 48530-48561 Da), further confirming the co-extraction of globulins in the albumin fraction, likewise observed through SDS-PAGE. Phaseolin is a diverse glycoprotein family with variable post-translational modifications and/or bound carbohydrates that may contribute to the three α-

phaseolins identified in the present work (Montoya et al., 2010). Lectins and lectin-like proteins (phytohemaglutinnin, α-amylase inhibitors), known as stress/defense proteins with carbohydrate-binding abilities, were also major protein species (Moreno et al., 1990). Lectins are typically considered to be albumins (Boye et al., 2010), but as previously noted, due to the sequential nature of the fractionation, unextracted albumins from the primary extraction (water) could be solubilized in subsequent aqueous media (saline and alkaline).

The prolamin fraction was unique in that the protein of highest abundance was an albumin that was not present in any of the other fractions. Yang et al. (2021) conducted a proteomic analysis of crude prolamins from kidney beans and similarly found a wide range of bean proteins in this fraction (e.g., legumin, lipoxygenase, cupin type-1 domain-containing protein, etc.). This further emphasizes that protein fractionation based on solubility does not yield pure fractions, potentially due to the slight degree of solubility of many bean proteins in all extraction media used.

Among the 180 identified proteins in the bean protein fractions, there were 20, 4, 2, and 6 proteins unique to the albumin, globulin, prolamin, and glutelin fractions, respectively (Supplementary Material Table S2.1). Those proteins were mostly identified as enzymes involved in metabolism, and were essentially negligible in peak area when compared to the more abundant storage proteins (e.g., phaseolin, legumin, albumin). However, proteomics profiling does not provide quantification in absolute terms; hence, future studies to quantify the ratios of various protein species in each fraction should be performed to illuminate possible quantitative differences in the extraction efficiency of each protein class.

Table 2.1. Top 10% identified proteins in the protein classes, AEP extract, and EAEP extract as determined by proteomic analysis with LC-MS/MS. Check marks (✓) denote that the protein is within the top 10% of identified proteins in the sample with respect to peak areas, while dots (•) denote that the protein is present, but not within the top 10%.

Description	Accession	-10lgP	Avg. Mass (Da)	ALB	GLO	PRO	GLU	AEP	EAEP
Storage Proteins									
Alpha-phaseolin	tr|X5CN36|X5CN36_PHAVU	277.73	48530	✓	✓	•	✓	✓	✓
Alpha-phaseolin	tr|X5D5D7|X5D5D7_PHAVU	278.91	48547	✓	✓	•	✓	✓	✓
Alpha-phaseolin	tr|X5CHW3|X5CHW3_PHAVU	279.55	48561	✓	✓	•	✓	✓	✓
Albumin_l_a domain-containing protein	tr|V7AJL4|V7AJL4_PHAVU	44.92	13961			✓			
Albumin-2	tr|F8QXP8|F8QXP8_PHAVU	173.53	25449	✓	✓		•	✓	•
Legumin	tr|F8QXP7|F8QXP7_PHAVU	266.13	68724	✓	✓		✓	✓	✓
Cupin type-1 domain-containing protein	tr|V7CFI9|V7CFI9_PHAVU	186.49	85002	✓	✓		✓	✓	•
Stress Related Proteins									
Lectin	tr|Q8RVX5|Q8RVX5_PHAVU	182.49	29569	✓	✓		✓	✓	✓
Antifungal lectin PVAP (Fragment)	sp|P84869|PVAP_PHAVU	64.69	1303	✓	✓		•	✓	✓
GNL-2 alpha subunit (Fragment)	tr|Q9S8B2|Q9S8B2_PHAVU	61.34	2407	✓	✓		✓	✓	•
Phytohemagglutinin	tr|Q8RVH2|Q8RVH2_PHAVU	170.9	29332	✓	✓	•	✓	✓	✓

Erythroagglutinating phytohemagglutinin	tr|V5QN77|V5QN77_PHAVU	216.42	29775	✓	✓	•	✓	✓	✓
Arcelin-4	sp|Q43629|ARC4_PHAVU	89.84	29451	✓	•		•	✓	
Arcelin 4-II	tr|Q8RVY3|Q8RVY3_PHAVU	90.02	30086	✓	•		•	✓	
Alpha amylase inhibitor-1	tr|A0T2V3|A0T2V3_PHAVU	181.78	27246	✓	✓		✓	✓	✓
Alpha-amylase inhibitor beta subunit, PHA-I beta subunit	tr|Q9S9E1|Q9S9E1_PHAVU	170.77	15404	✓	✓		•	✓	✓
Alpha-amylase inhibitor like protein	tr|Q9SMH0|Q9SMH0_PHAVU	136.07	28900	✓	•		•	✓	
Group 3 late embryogenesis abundant protein	tr|Q2N1E0|Q2N1E0_PHAVU	163.84	50640	•	•		•	✓	
Metabolism Related Proteins									
GH18 domain-containing protein	tr|V7AIB2|V7AIB2_PHAVU	99.27	35245	✓	•			•	•
Alcohol dehydrogenase 1	tr|T2DLR9|T2DLR9_PHAVU	144.37	41355	✓			✓	•	•
Purple acid phosphatase	tr|V7AEU9|V7AEU9_PHAVU	154.25	45249	•	✓		✓	•	
Peptidase A1 domain-containing protein	tr|V7BPV5|V7BPV5_PHAVU	133.89	46270	•	•		✓	✓	✓
Fe(3+)-Zn(2+) purple acid phosphatase	sp|P80366|PPAF_PHAVU	162.45	52857	•	✓		✓	•	
Lipoxygenase	tr|V7BX14|V7BX14_PHAVU	187.99	97360	✓	✓		✓	✓	
Lipoxygenase	tr|V7BZK0|V7BZK0_PHAVU	209.55	97545	✓	✓		✓	✓	•
Alpha-1 4 glucan phosphorylase	tr|V7C329|V7C329_PHAVU	171.13	111076	✓	•		•	•	•

Other									
Uncharacterized protein	tr|V7B7H6|V7B7H6_PHAVU	149.86	22923	✓	•		✓	•	
Uncharacterized protein	tr|V7BIT8|V7BIT8_PHAVU	87.06	25290	•	•		✓	•	
Uncharacterized protein	tr|V7BBR4|V7BBR4_PHAVU	155.44	26165	•	•		•	✓	
Uncharacterized protein	tr|V7C790|V7C790_PHAVU	106.22	26380	•			✓	✓	
Uncharacterized protein	tr|V7CL08|V7CL08_PHAVU	154.13	29200	•	✓		✓	✓	•
Uncharacterized protein	tr|V7BN55|V7BN55_PHAVU	206.4	33641	✓	✓		✓	✓	✓
Uncharacterized protein	tr|V7BFT4|V7BFT4_PHAVU	150.65	38845	✓	•		✓	•	•
Uncharacterized protein	tr|V7BFL3|V7BFL3_PHAVU	207.77	57274	✓	✓	•	✓	✓	✓

-10lgP: Score from Peaks Xpro software

ALB: albumin-rich fraction, GLO: globulin-rich fraction, PRO: prolamin-rich fraction, GLU: glutelin-rich fraction

Check marks (✓) denote that the protein is within the top 10% of identified proteins in the sample (by summed area), while dots (•) denote that the protein is present, but not within the top 10%.

Glycoprofiling of subcritical water extracts

The subcritical water extraction of insoluble byproducts remaining after enzyme-assisted aqueous extraction process (EAEP) was investigated, with a focus on the identification and profiling of oligosaccharides. While there are a few studies of legume byproducts extracted by subcritical water (Ramirez et al., 2021; Wiboonsirikul et al., 2013), the use of subcritical water extraction to release oligosaccharides from black bean waste following protease treatment is wholly novel.

For building a library, all fractions were normalized to their respective total soluble carbohydrate content as determined by the phenol-sulfuric acid method (DuBois et al., 1956), ensuring that roughly the same amount of carbohydrates were analyzed by LC-MS/MS for all fractions. This normalization ensured that oligosaccharides would be adequately detected among all fractions, enabling the inclusion of a maximum number of oligosaccharides in the library. All oligosaccharides in the library consisted of Hexose (Hex), Pentose (Pent), deoxyhexose (dHex), and acidic sugars (HexA) as all possible constituent monosaccharides. Possible hexoses include glucose, galactose, and mannose; pentoses include xylose; deoxyhexoses include fucose (the primary monosaccharides of fucoidan, the polysaccharide of interest); and acidic sugars include galacturonic acid and glucuronic acid, among others. All features in the bean library are reported in Supplementary Table S2.2.

Among all subcritical temperatures and fractions, a total of 173 oligosaccharides were identified and included in the library as reported in Figure 2.4. Oligosaccharides were classified by their monosaccharide composition as containing hexoses-only, pentose-only, hexose-pentose, and remaining others containing dHex and/or HexA. The reported results consist of manually summed in-source fragments and dimer and trimer aggregates in order to approximate the true abundance of each oligosaccharide (Huang et al., 2022). Figure 2.5 reports the total counts of

oligosaccharides for each subcritical water extraction fraction and temperature as classified by their monosaccharide composition. Results are presented through this classification to assess the differences in oligosaccharide profiles that stem from various subcritical water extraction temperatures.

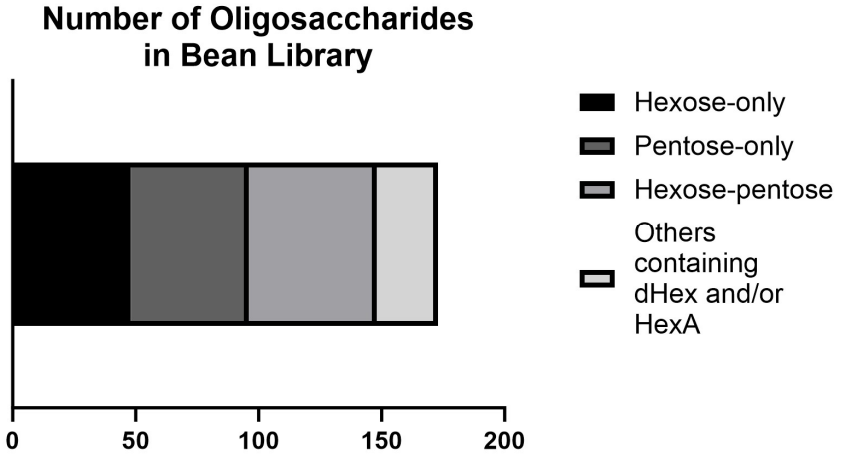


Figure 2.4. Total number of oligosaccharides in the bean library, categorized based on constituent monosaccharides. Oligosaccharides were categorized as containing only hexoses, only pentoses, hexoses and pentoses, and all others containing deoxyhexoses (dHex) and/or acidic sugars (HexA).

Figure 2.5 reports the total counts of oligosaccharides identified in subcritical water extracts. Results demonstrated that varied extraction temperatures are reflected in each condition's glycoprofile. The highest number of identified oligosaccharides was observed under subcritical conditions at 200 °C, while 120 °C exhibited slightly over half of the total number compared to 200 °C. Subcritical water extraction at 160 °C showed a minor decrease in the total number of oligosaccharides compared to 200 °C but still yielded more than the lowest temperature. These findings are consistent with studies of wheat and rice bran which reported that carbohydrate content increased with increasing temperature up to 180-200 °C (Chiou et al., 2011; Kataoka et al., 2008; Wiboonsirikul et al., 2007).

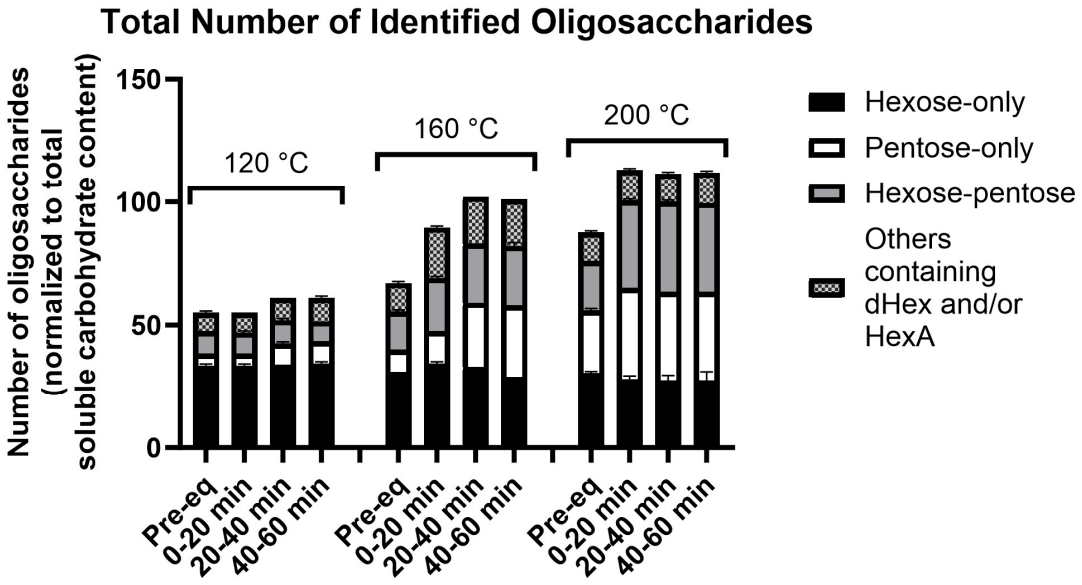


Figure 2.5. Total counts of oligosaccharides identified in subcritical water extracts. Extracts were normalized to soluble carbohydrate content for analysis. Subcritical water extraction was evaluated at three temperatures, 120 °C, 160 °C, and 200 °C. The total time of extraction was 60 min, with fractions collected during pre-equilibrium (the time for the extraction vessel to achieve set pressure), 0 - 20 min, 20 - 40 min, and 40- 60 min. At each temperature and time, oligosaccharides are categorized by their monosaccharide composition.

Approximately 28% of the identified oligosaccharides consisted solely of hexoses (hexose-only oligosaccharides) (Fig. 2.4). Among this group, 120 °C consistently yielded the highest number of identified oligosaccharides (34-35) across all temperatures and extraction times (Fig. 2.5). At 160 °C, earlier subcritical water extraction times (0-20 min and 20-40 min) showed comparable numbers to 120 °C, but the pre-equilibrium and final subcritical water extraction fraction (40-60 min) exhibited a slight decrease in the number of hexose-only oligosaccharides compared to 120 °C. The highest temperature (200 °C) resulted in the lowest number of hexose-only oligosaccharides, with the pre-equilibrium fraction having the highest abundance of oligosaccharides (31) compared to the later fractions at 200 °C (containing on average 28 oligosaccharides). This suggests that hexose-only oligosaccharides may be completely extracted during the pre-equilibrium fraction at 200 °C and are subsequently degraded at this high temperature with extended extraction times.

Pentose-only and hexose-pentose oligosaccharides accounted for approximately 28% and 30% respectively of all identified oligosaccharides (Fig. 2.4). This group exhibited a distinct profile from that of hexose-only oligosaccharides, displaying a nearly inverse profile. The data shows that pentose-only oligosaccharides are predominantly extracted at 200 °C, particularly within subcritical conditions (Fig. 2.5). A considerable fraction of these oligosaccharides were also identified at 160 °C, however only the later subcritical fractions (20-40 min and 40-60 min) exhibited values comparable to those at 200 °C, indicating a need for sufficient extraction time at this intermediate temperature. At the lowest temperature (120 °C), a sharp decrease in the number of pentose-only and hexose-pentose oligosaccharides was observed, suggesting that either 120 °C is an insufficient temperature for the release of pentose-only oligosaccharides or that 60 min of subcritical water extraction at this temperature is not long enough to adequately release these oligosaccharides.

The remaining oligosaccharides which contained dHex and/or HexA constituted a smaller proportion of the total identified oligosaccharides (roughly 14%) (Fig. 2.1), but displayed a unique profile when compared to all other groups. The profile of these oligosaccharides resembles the hexose-only group for the highest and intermediate temperatures, with a moderate number detected in all fractions at 200 °C and the pre-equilibrium fraction at 160 °C (Fig. 2.5). Additionally, a much greater number of oligosaccharides in this group were detected in the subcritical fractions of 160 °C. However, unlike hexose-only oligosaccharides, there was a steep decline in the number of detected oligosaccharides with decreasing extraction temperature, particularly in the 120 °C samples. This suggested that, similarly to pentose-only and hexose-pentose oligosaccharides groups, subcritical water extraction at 120 °C for 60 min does not sufficiently release oligosaccharides containing dHex and/or HexA.

Figures 2.6 A-D depict glycoprofiles by heatmap. When visualizing glycoprofiling peak area results by heatmap, interesting profiles are observed depending on the extraction temperature and length of extraction. The heatmaps are categorized as hexose-only, pentose-only, hexose-pentose, and all other oligosaccharides containing dHex and HexA. Individual oligosaccharides are reported on each row of the heatmap, and the four-digit codes along the y-axis indicate the number of monosaccharides making up the oligosaccharide (Hex_dHex_Pent_HexA). All reported values were normalized to total soluble carbohydrate concentration as determined by the phenol-sulfuric acid method (DuBois et al., 1956) prior to SPE and analysis. Thus, the heatmaps show the oligosaccharides profile per roughly the same amount of carbohydrate for all samples.

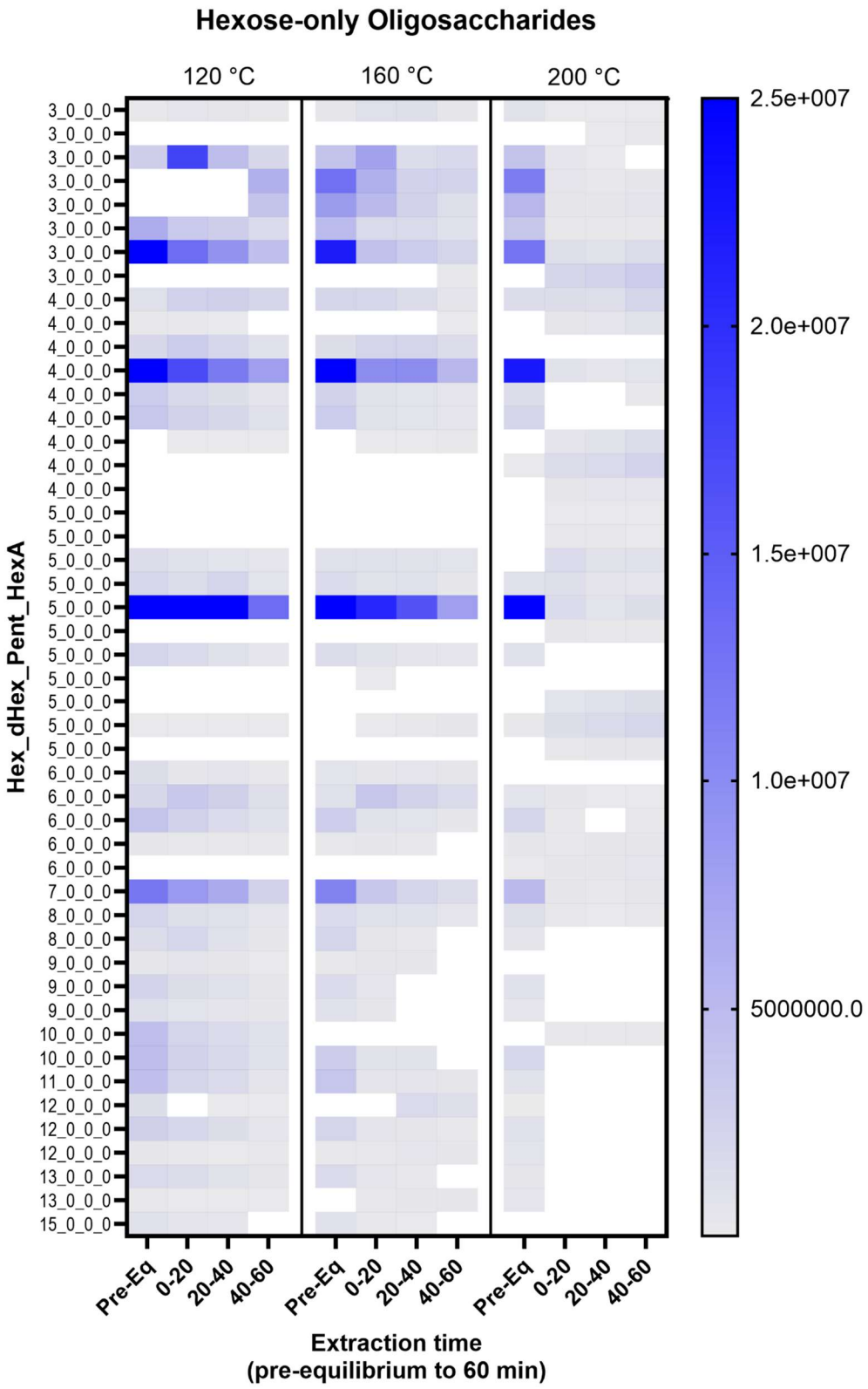
Peak area results of oligosaccharides containing hexose-only and any containing dHex and HexA revealed a clear time-temperature extraction relationship (Fig. 2.6 A, D). At high temperature (200 °C), the majority of hexose-only oligosaccharides were readily extracted during the pre-equilibrium fraction. Very few hexose-only oligosaccharides with higher degrees of polymerization (DP), were detected in subsequent fractions. As extraction temperature decreased, the extraction of hexose-only oligosaccharides became more evenly distributed across the subsequent fractions. This is particularly evident for the lowest extraction temperature (120 °C) where a consistent response was observed among all extraction fractions from pre-equilibrium to the final 40-60 minute fraction, although the response slightly decreased over time.

Regarding pentose-only and hexose-pentose oligosaccharides (Fig. 2.6 B,C), nearly all were identified in the samples extracted at 200 °C. Similar to the hexose-only oligosaccharides, a few oligosaccharides were readily extracted during the pre-equilibrium stage but were not detected in subsequent fractions. However, most pentose-only and hexose-pentose oligosaccharides

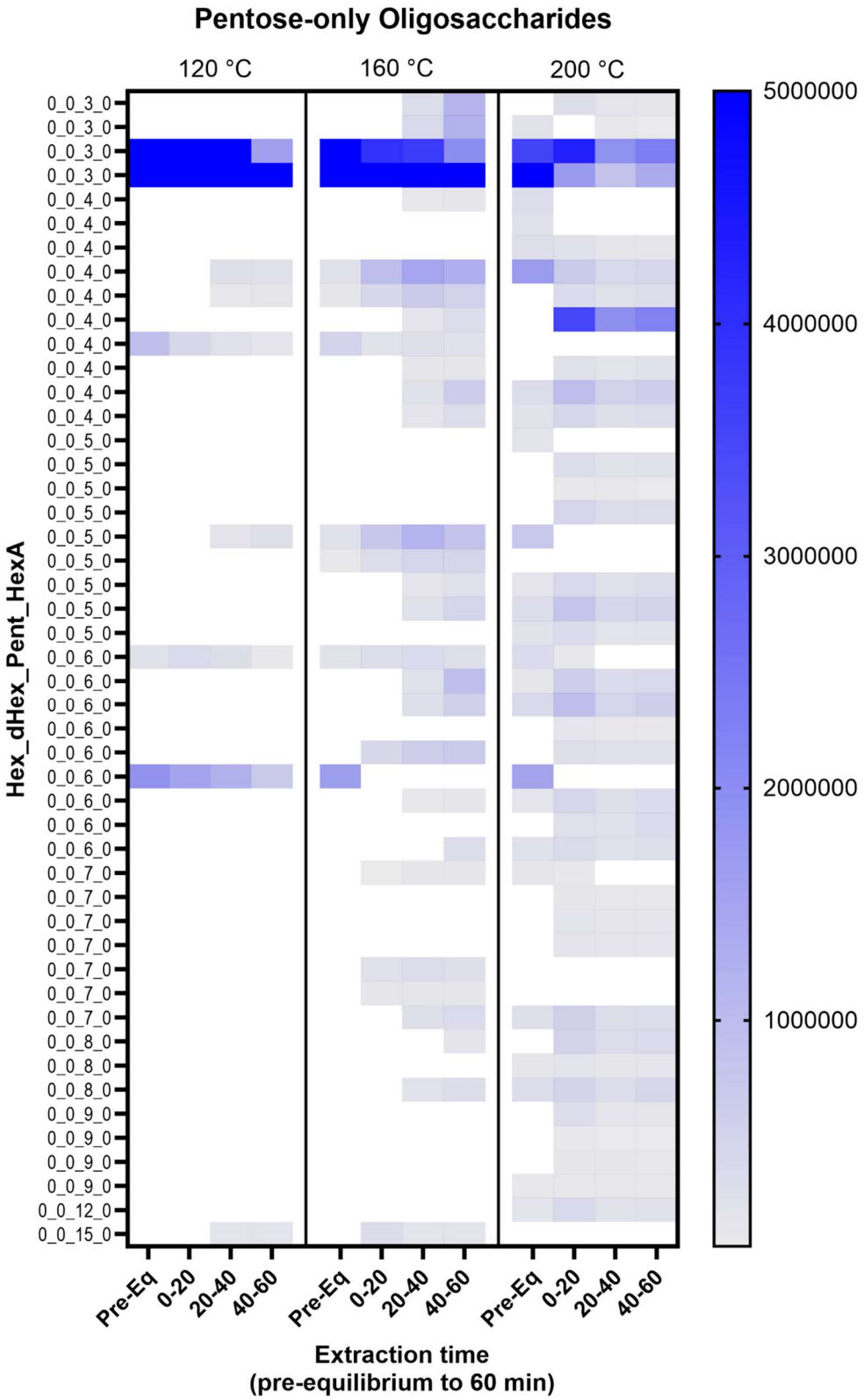
were identified throughout the entire extraction process, from pre-equilibrium to 60 minutes. Some of these oligosaccharides were exclusively detected between 0-60 minutes, indicating the important role of subcritical extraction conditions for their release. At 160 °C, the detected pentose-only and hexose-pentose oligosaccharides were mainly associated with later fractions (20-40 minutes and 40-60 minutes), highlighting the importance of longer extraction times at lower extraction temperatures. At the lowest temperature (120 °C), a very small fraction of these oligosaccharides were detected.

With the aim of normalizing fractions by total carbohydrate concentration to create the library, significant efforts were made to dilute or concentrate the samples to achieve equal carbohydrate levels across all fractions. Phenol-sulfuric assay results showed orders of magnitudes in concentration variations between the pre-equilibrium fraction and the other subcritical fractions, with the pre-equilibrium fraction consistently exhibiting much higher carbohydrate concentration. This is consistent with results by Wiboonsirikul et al. (2013) which found in subcritical water extraction of okara, heating at any treatment temperatures for longer than 5 min resulted in decreased carbohydrate content. In order to provide a more accurate representation of the subcritical water extraction as they were generated, peak abundances were reported without normalization and presented in the unnormalized heatmaps provided in Supplementary Figures S2.1 A-D. The unnormalized data shows that if an oligosaccharide was detected in the pre-equilibrium fraction, this fraction consistently displayed the highest abundance.

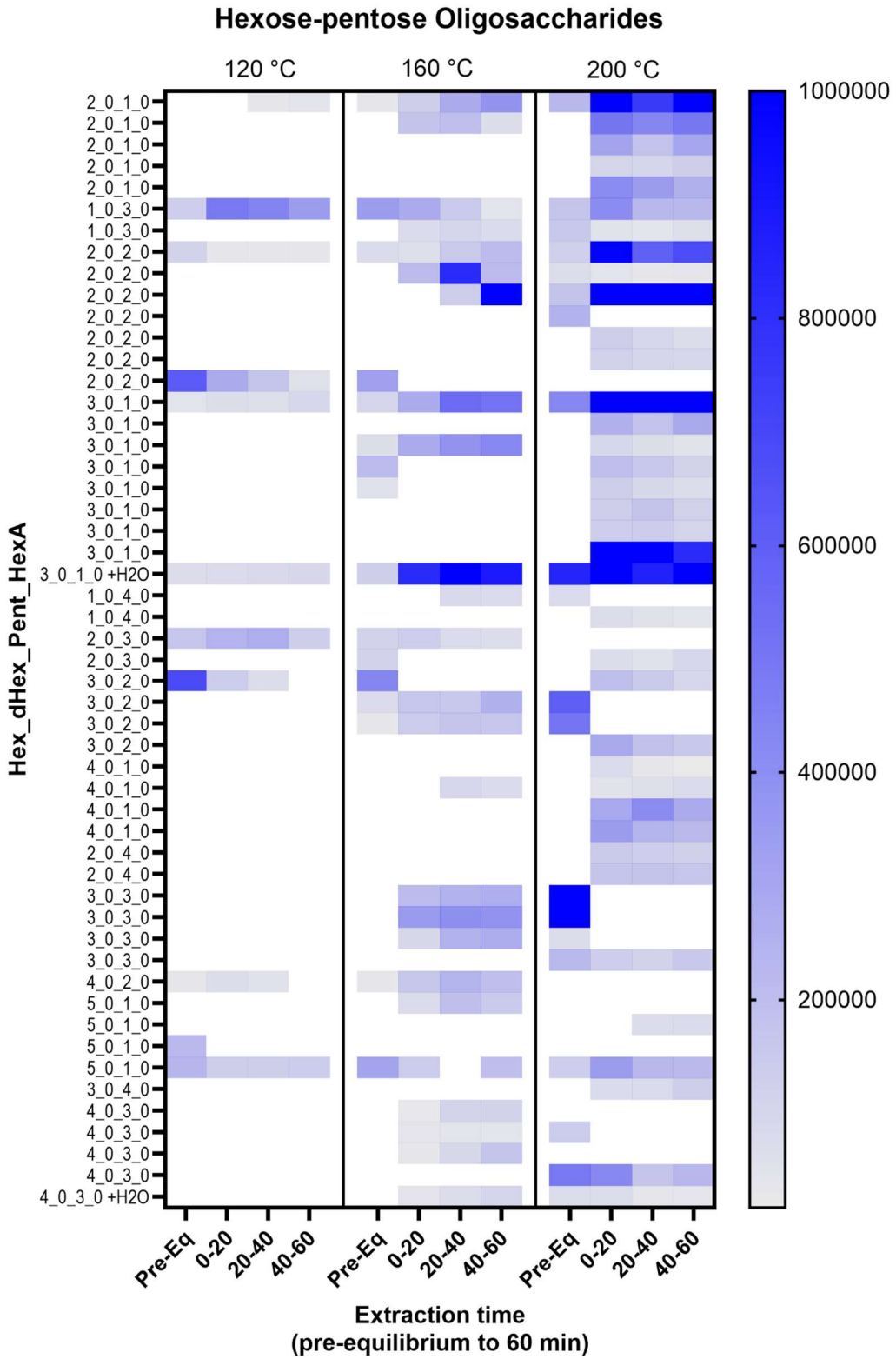
A)


Hexose-only Oligosaccharides
120 °C
160 °C
200 °C
Hex_dHex_Pent_HexA
3_0_0_0
4_0_0_0
5_0_0_0
6_0_0_0
7_0_0_0
8_0_0_0
9_0_0_0
10_0_0_0
11_0_0_0
12_0_0_0
13_0_0_0
15_0_0_0
Pre-Eq
0-20
20-40
40-60
Extraction time
(pre-equilibrium to 60 min)
2.5e+007
2.0e+007
1.5e+007
1.0e+007
5000000.0

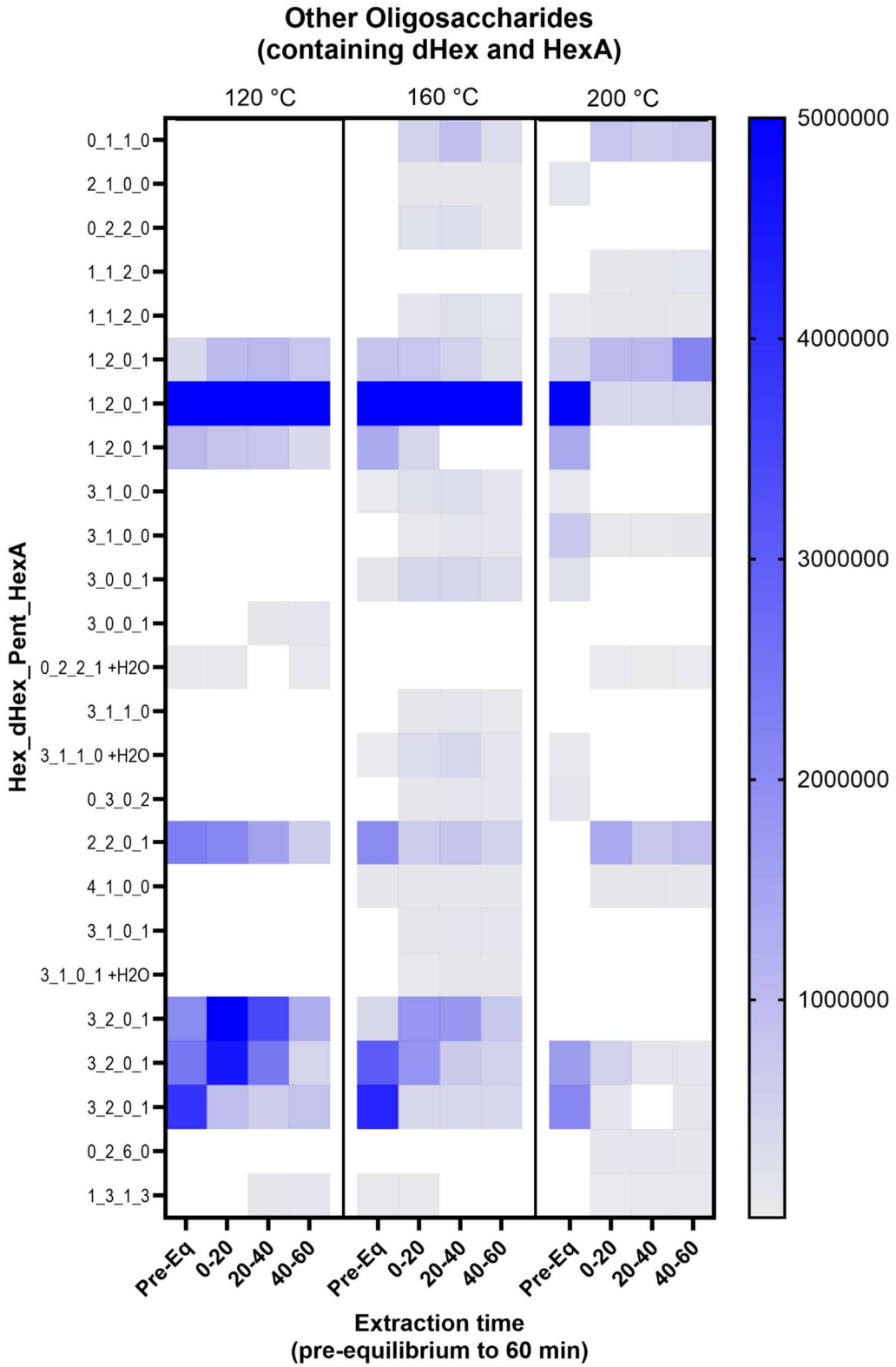
B)


Pentose-only Oligosaccharides
120 °C
160 °C
200 °C
Hex_dHex_Pent_HexA
0_0_3_0
0_0_4_0
0_0_5_0
0_0_6_0
0_0_7_0
0_0_8_0
0_0_9_0
0_0_12_0
0_0_15_0
5000000
4000000
3000000
2000000
1000000
Pre-Eq
0-20
20-40
40-60
Extraction time
(pre-equilibrium to 60 min)

C)


Hexose-pentose Oligosaccharides
120 °C
160 °C
200 °C
Hex_dHex_Pent_HexA
2_0_1_0
2_0_1_0
2_0_1_0
2_0_1_0
2_0_1_0
1_0_3_0
1_0_3_0
2_0_2_0
2_0_2_0
2_0_2_0
2_0_2_0
2_0_2_0
2_0_2_0
2_0_2_0
3_0_1_0
3_0_1_0
3_0_1_0
3_0_1_0
3_0_1_0
3_0_1_0
3_0_1_0
3_0_1_0
3_0_1_0 +H2O
1_0_4_0
1_0_4_0
2_0_3_0
2_0_3_0
3_0_2_0
3_0_2_0
3_0_2_0
3_0_2_0
4_0_1_0
4_0_1_0
4_0_1_0
4_0_1_0
2_0_4_0
2_0_4_0
3_0_3_0
3_0_3_0
3_0_3_0
3_0_3_0
4_0_2_0
5_0_1_0
5_0_1_0
5_0_1_0
5_0_1_0
3_0_4_0
4_0_3_0
4_0_3_0
4_0_3_0
4_0_3_0
4_0_3_0 +H2O
Pre-Eq
0-20
20-40
40-60
Extraction time
(pre-equilibrium to 60 min)
1000000
800000
600000
400000
200000

D)


Other Oligosaccharides
(containing dHex and HexA)
120 °C
160 °C
200 °C
Hex_dHex_Pent_HexA
0_1_1_0
2_1_0_0
0_2_2_0
1_1_2_0
1_1_2_0
1_2_0_1
1_2_0_1
1_2_0_1
3_1_0_0
3_1_0_0
3_0_0_1
3_0_0_1
0_2_2_1 +H2O
3_1_1_0
3_1_1_0 +H2O
0_3_0_2
2_2_0_1
4_1_0_0
3_1_0_1
3_1_0_1 +H2O
3_2_0_1
3_2_0_1
3_2_0_1
0_2_6_0
1_3_1_3
Pre-Eq
0-20
20-40
40-60
Extraction time
(pre-equilibrium to 60 min)
5000000
4000000
3000000
2000000
1000000

Figures 2.6. Glycoprofile heatmaps of subcritical water extractions of bean insoluble, with oligosaccharides categorized as containing hexose-only (A), pentose-only (B), hexose-pentose (C), and all others containing dHex and/or HexA (D). Individual oligosaccharides are reported on each row of the heatmaps, and the four-digit codes along the left y-axis indicate the number of monosaccharides making up the oligosaccharide (Hex_dHex_Pent_HexA). The right color legend indicates peak areas. All samples were analyzed after normalization to soluble carbohydrate content. Oligosaccharides for each subcritical temperature (120 °C, 160 °C, 200 °C) and extraction time (pre-equilibrium, 0 - 20 min, 20 - 40 min, and 40 - 60 min) are reported.

Discussion

Bean proteins (AEP, EAEP, Osborne fractionation)

The albumin-rich fraction exhibited the highest protein extractability although it is widely reported that globulins are the primary protein class in beans (50-60% of total protein) (Ma and Harwalkar, 1984). This discrepancy may be attributed to the high solubility (~80%) of phaseolin (the major globulin in common beans) in water due to the native ionic strength of the bean flour (Shpande and Nielsen, 1987). Cross-contamination of bean albumins and globulins in sequential fractionation has also been reported by Deshpande and Nielsen (Shpande and Nielsen, 1987), who similarly observed higher protein yields in the water-soluble fraction (52.4% of total protein) than the salt-soluble fraction (37.9% of total protein) for black beans.

In the EAEP, the use of alkaline protease significantly improved protein extraction yields. Increased extractability with enzyme-assisted extraction has been attributed to proteolysis, which solubilizes previously insoluble proteins and releases proteins entrapped within the complex cell matrix (Campbell and Glatz, 2009; de Moura et al., 2008; Souza Almeida et al., 2021; Souza et al., 2019). Scanning electron microscopy imaging of navy bean flour has shown the presence of proteinaceous material associated with the starch granules in the flour (Berg et al., 2012), suggesting that protease action in the EAEP may have facilitated the release of starch-bound or cell-wall-bound proteins, therefore enhancing protein extractability.

As expected, proteomics showed many overlapping proteins among Osborne fractions due to the sequential nature of extraction. More notably, proteomic results revealed differences between the protein composition between the AEP and EAEP extracts, confirming that proteolysis altered the protein composition of the fraction. Specifically, proteins such as arcelin (a carbohydrate-binding protein) and some metabolism-related proteins (purple acid phosphatase, lipoxygenase) were not detected in the EAEP extract, suggesting complete hydrolysis of those proteins by alkaline protease.

Somewhat surprisingly, the number of identified proteins was lower in the EAEP extract compared to the AEP extract. This is likely attributed to multiple factors. Firstly, following EAEP, highly variable length peptides were released depending on the degree of hydrolysis. Due to the conventional bottom-up proteomics workflow utilized, any peptides released through EAEP were removed with the supernatant in the initial protein precipitation steps. Secondly, the non-specific nature of the protease used in EAEP led to cleavage at unpredictable sites within proteins. In contrast, proteomics analysis typically utilizes enzymes with specific cleavage sites, such as trypsin, to simplify data processing and database searching. Without the knowledge of which amino acids could have been cleaved by the protease, even proteins that were only partially hydrolyzed and remained in solution following the initial protein precipitation may not have sufficient peptide sequences to be adequately identified by database searching. Consequently, the lack of knowledge regarding the cleavage sites of the protease in the EAEP extracted protein limited the number of adequately sequenced peptides, therefore artificially reducing the number of protein identifications compared to AEP extracts.

It should be noted that the higher degree of hydrolysis for EAEP did not always result in a complete loss of protein identification. This is evident with phaseolin, which despite being predominantly hydrolyzed according to SDS-PAGE analysis, remained one of the major

identified proteins in the EAEP extract. This is likely due to phaseolin's abundance as a major storage protein. Even partially hydrolyzed phaseolin in EAEP was abundant enough to be sequenced and identified by database searching. For this reason, proteomics data should be considered in conjunction with other methods of protein analysis, such as protein visualization by SDS-PAGE, to accurately assess the extent of hydrolysis for highly abundant proteins.

This study aimed to investigate the impact of different extraction methods on the extraction of protein classes in black beans. For this aim, sequentially-extracted protein fractions (albumin-, globulin-, prolamin-, and glutelin-rich) were analyzed along with aqueous extraction process (AEP) and enzyme-assisted aqueous extraction process (EAEP). SDS-PAGE and proteomics analysis revealed minor variations in protein compositions/profiles among Osborne fractions, with a significant overlap of proteins across different fractions. Sequential fractionation was found to be useful in understanding the extraction of common bean proteins, despite not providing distinct separation of protein classes. The use of an alkaline protease significantly increased protein extractability compared to aqueous extraction alone. The application of proteomics techniques allowed for the identification of proteins that were susceptible to proteolysis, thus providing valuable insights into the impact of enzyme-assisted extraction. These findings underscore the significance of enzyme-assisted extraction and the utility of proteomics in elucidating variations in protein composition resulting from different extraction methods. Such insights are crucial in guiding the development of environmentally friendly extraction approaches and judicious utilization of natural resources.

Bean insoluble subcritical water extracts glycoprofiling

Glycomics analysis offered valuable insights into the impacts of extraction temperature and time on the glycoprofile of extracted insoluble bean byproducts. By categorizing oligosaccharides according to their constituent monosaccharides (hexose-only, pentose-only, hexose-pentose,

others containing dHex and/or HexA), distinct glycoprofiles are observed across different extraction temperatures and time.

In particular, temperature had a significant impact on the number and composition of identified oligosaccharides. Higher temperatures generally led to a greater number of identified oligosaccharides as has been found in other studies (Chiou et al., 2011; Kataoka et al., 2008; Wiboonsirikul et al., 2007), but glycomics analysis revealed that the distribution and composition of the oligosaccharides varied depending on the temperature and extraction time. Specifically, at the highest temperature (200 °C), pentose-only and hexose-pentose oligosaccharides were readily extracted, whereas at the lowest temperature (120 °C), a very small fraction of these oligosaccharides were detected. This suggests that there is a minimum temperature threshold that is necessary for adequate extraction of these oligosaccharides. Conversely, at the lowest temperature (120 °C), most hexose-only oligosaccharides were detected. These hexose-only oligosaccharides decreased as temperature increased, with the highest temperature (200 °C) showing the lowest number and abundance of these oligosaccharides. This may indicate potential degradation of hexose-only oligosaccharides at higher temperatures. Further investigations quantifying monosaccharides in these extracts may reveal whether or not these hexose-only oligosaccharides are being degraded into their constituent monosaccharides at high temperatures.

The role of extraction time was also evident, with clear time-temperature relationships observed. Longer extraction times were required at lower extraction temperatures to extract the same oligosaccharides. However, when considering data that was not normalized by total soluble carbohydrate content, pre-equilibrium fractions consistently exhibited the most abundant and diverse glycoprofiles. This suggests that achieving a high temperature is of primary importance

for releasing oligosaccharides, and extended extraction times may not be necessary, particularly considering the low yields in later fractions.

It is important to acknowledge that the results used to create a library were normalized, ensuring that roughly the same amount of carbohydrates were used for SPE and nanoLC-QToF analysis. While this approach is useful for generating novel oligosaccharide libraries and obtaining the maximum number of identified oligosaccharides generated from subcritical water extraction it may not be fully representative of the extraction itself or any potential industrial applications. Notably, the concentrations of total soluble carbohydrates in the later fractions (20-40 minutes and 40-60 minutes) were orders of magnitude lower compared to the pre-equilibrium fractions. To achieve normalized results, many additional concentration steps were needed for later subcritical fractions, while pre-equilibrium samples required dilution.

To obtain more representative results for glycoprofiling studies, analyzing extracts as they are generated without additional manipulation may be a truer representation of the extraction process and potentially offers insights that align better with industrial applications.
In particular, unnormalized data revealed that if an oligosaccharide was detected in the pre-equilibrium fraction, this fraction consistently exhibited the highest abundance. This observation is crucial when considering the practical application of this research in industry. It suggests that a one-hour extraction under subcritical conditions may not be necessary, as the majority of oligosaccharides are likely released when sufficiently high temperatures are reached. However, prior to selecting optimal extraction conditions, it is important to first determine the biological characteristics of the obtained extractions and select for desired biological properties to guide extraction.

This work demonstrates how glycomics techniques can provide insight to the effects of time and temperature of subcritical water extraction. Given that unique profiles were observed depending on the extraction temperature, these results show that there are opportunities for optimization. It is possible to adjust extraction conditions to preferentially extract oligosaccharides of interest, taking into account their structure and composition. Furthermore, the construction and utilization of a novel oligosaccharide library to assess the products that come from the breakdown of insoluble lentil byproducts during subcritical water extraction highlight the value of this green extraction technique and the valuable insights provided by glycomics techniques to guide the advancement of these new environmentally friendly extraction approaches.

Conclusion

Proteomics was employed for the identification of proteins in black beans, focusing on the impact of different extraction media and proteolysis that results from EAEP. For Osborne fractions, proteomics analysis revealed similarities in protein profiles among the albumin, globulin, and glutelin fractions, while the prolamin fraction exhibited distinct differences. Major proteins, such as α-phaseolin, phytohemagglutinin, and erythroagglutinating phytohemagglutinin, were detected in all protein fractions, demonstrating the continued extraction of proteins during sequential fractionation. The use of alkaline protease in EAEP increased protein extraction yields by hydrolyzing previously insoluble proteins and releasing peptides and proteins trapped within the cell matrix. Notably, proteomics analysis revealed differences in protein composition between AEP and EAEP extracts, indicating the influence of proteolysis on specific proteins and overall protein composition. The number of identified proteins was lower in the EAEP extract, possibly due to variable peptide lengths released by the protease and the non-specific cleavage sites of the protease used. Nevertheless, highly abundant proteins like phaseolin remained detectable and identifiable even when partially hydrolyzed. It is important to integrate proteomics data with other protein analysis methods,

such as SDS-PAGE, to accurately assess the extent of hydrolysis, especially for abundant proteins. Overall, this work provides valuable insights into protein extraction methods and their impact on protein composition in black beans, contributing to the understanding of protein characterization in legume crops and their potential application in industry.

Glycomics revealed the significant impact of extraction temperature and time on glycoprofiles of the subcritical water extraction of insoluble bean byproducts. Higher temperatures generally resulted in a greater number of identified oligosaccharides, with distinct profiles observed for different temperature conditions. Pentose-only and hexose-pentose oligosaccharides were predominantly extracted at higher temperatures, whereas hexose-only oligosaccharides exhibited higher abundance at lower temperatures. This suggested the presence of a minimum temperature threshold for the extraction of certain oligosaccharides. Extraction time also played a role, with longer times required at lower temperatures. However, pre-equilibrium fractions consistently displayed the most abundant glycoprofiles particularly at higher temperatures, indicating the importance of achieving a high temperature for effective oligosaccharides release. Utilizing results normalized by total carbohydrate content allowed for the generation of a comprehensive library, however it is noted that this may not fully represent the extraction process. As a result, the accuracy of identifying prospective industrial uses for the extracts could be influenced. The findings highlight the need for careful selection of temperature and extraction time to obtain desired oligosaccharides compositions and offer insights for optimization in subcritical water extraction processes. Further investigations, including quantification of monosaccharides, could provide additional understanding of oligosaccharides degradation and release mechanisms at different temperatures. Furthermore, future evaluations of the biological properties of the extracted oligosaccharides are warranted to guide the selection of best processing conditions. Overall, glycomics techniques provide valuable tools for studying the

effects of time and temperature in subcritical water extraction and help to guide the development of green and effective extraction methods.

Acknowledgements

This research was supported by funding by the USDA Pulse Crop Health Initiative (grant number 58-3060-0-044-0).

Supplementary materials

Supplementary Table S2.1. All black bean proteins identified through proteomics analysis. Protein description, accession, average mass, PeaksXPro quality score (-10lgP), posttranslational modifications (PTM), and peak areas are reported. Peak areas are separated into the ALB = albumin-rich, GLO = globulin-rich, PRO = prolamin-rich, and GLU = glutelin-rich fractions, and the aqueous extraction process (AEP), and enzyme-assisted aqueous extraction process (EAEP). Peak areas are reported as the sum of the top three peptides.

Description	Accession	Avg. Mass (Da)	-10 lgP	PTM	Area ALB	Area GLO	Area PRO	Area GLU	Area AEP	Area EAEP
Chain A phaseolin beta-type precursor	pdb|1PHS|A	44992	276	Deamidation (NQ)	1.67E+07	3.88E+07	1.10E+04	6.04E+06	3.02E+07	3.83E+07
alpha-phaseolin	AHW49421.1	48561	271	Deamidation (NQ); Dioxidation (M)	1.58E+07	3.79E+07	5.13E+03	4.89E+06	2.81E+07	3.90E+07
alpha-phaseolin	AHW49417.1	48530	268	Deamidation (NQ); Dioxidation (M)	1.48E+07	3.77E+07	1.08E+03	3.68E+06	2.67E+07	3.73E+07
legumin	ADR30064.1	68724	255	Carbamidomethylation; Deamidation (NQ); Phosphorylation (STY)	2.55E+06	4.79E+06	0.00E+00	3.60E+06	5.00E+06	2.62E+06
alpha-phaseolin	AHW49409.1	48789	229	Deamidation (NQ); Dioxidation (M)	8.48E+06	2.24E+07	5.13E+03	3.64E+06	1.71E+07	2.99E+07
Chain B Erythroagglutinin	pdb|3WCR|B	27599	211	Deamidation (NQ)	5.79E+06	3.68E+06	9.62E+01	3.77E+06	6.54E+06	5.65E+05
hypothetical protein PHAVU_007G173600g	XP_007144650.1	57274	207	Carbamidomethylation; Deamidation (NQ)	6.65E+05	1.15E+06	2.34E+02	1.38E+06	1.07E+06	1.43E+06
hypothetical protein PHAVU_005G156700g	XP_007150486.1	97545	201	Carbamidomethylation; Deamidation (NQ)	7.53E+05	5.25E+05	0.00E+00	1.55E+05	1.97E+05	1.55E+03

hypothetical protein PHAVU_006G120100g	XP_007147388.1	33641	199		3.36E+05	1.38E+06	0.00E+00	9.78E+05	5.17E+05	1.68E+06
hypothetical protein PHAVU_002G027900g	ESW28909.1	85002	193	Carbamidomethylation; Deamidation (NQ)	5.18E+05	2.66E+05	0.00E+00	5.89E+05	4.80E+05	6.81E+03
hypothetical protein PHAVU_005G157000g	ESW22484.1	97360	189	Carbamidomethylation; Dioxidation (M)	6.70E+05	5.24E+05	0.00E+00	1.73E+05	2.56E+05	0.00E+00
lectin	CAD29133.1	29569	171	Deamidation (NQ)	2.47E+06	1.08E+06	0.00E+00	1.04E+06	2.08E+06	1.28E+05
hypothetical protein PHAVU_004G034400g	XP_007151297.1	111076	169	Carbamidomethylation	1.82E+05	4.46E+03	0.00E+00	1.73E+04	1.26E+05	0.00E+00
alpha amylase inhibitor-1 precursor	ABK79078.1	27246	168	Deamidation (NQ)	2.38E+06	5.55E+05	0.00E+00	5.07E+05	1.61E+06	6.01E+05
phytohemagglutinin	CAD28674.1	29332	166	Deamidation (NQ)	2.50E+06	1.41E+06	9.62E+01	4.58E+05	2.87E+06	4.44E+05
albumin-2	ADR30065.1	25449	162	Carbamidomethylation; Deamidation (NQ)	1.34E+06	5.62E+05	0.00E+00	4.69E+05	8.70E+05	8.54E+04
hypothetical protein PHAVU_004G117100g	XP_007152289.1	141378	160	Phosphorylation (STY)	3.26E+04	3.82E+04	0.00E+00	1.37E+04	2.74E+04	0.00E+00
alpha-amylase inhibitor beta subunit PHA-I beta subunit	AAB50854.1	15404	157		1.74E+06	4.96E+05	0.00E+00	1.84E+05	1.30E+06	4.68E+05
hypothetical protein PHAVU_005G030100g	XP_007148975.1	36195	155		1.53E+04	3.68E+04	0.00E+00	1.31E+05	3.39E+04	6.36E+03
Chain D Phytohemagglutinin-l	pdb|1FAT|D	27419	154	Deamidation (NQ)	2.65E+06	1.64E+06	9.62E+01	6.61E+05	3.12E+06	5.67E+05

group 3 late embryogenesis abundant protein	ABA26579.1	50640	153	Carbamidomethylation; Deamidation (NQ); Phosphorylation (STY)	3.04E+04	6.15E+03	0.00E+00	2.21E+04	3.50E+05	0.00E+00
Fe(3+)-Zn(2+) purple acid phosphatase	sp|P80366.3|PPAF_PHAVU	52857	150	Carbamidomethylation	5.66E+04	8.32E+05	0.00E+00	2.27E+05	6.36E+04	0.00E+00
hypothetical protein PHAVU_011G008800g	XP_007131379.1	45249	147	Carbamidomethylation	7.80E+04	9.91E+05	0.00E+00	2.40E+05	4.52E+04	0.00E+00
hypothetical protein PHAVU_008G144500g	XP_007140814.1	22923	146	Deamidation (NQ)	1.65E+05	8.76E+04	0.00E+00	4.90E+05	7.30E+04	0.00E+00
hypothetical protein PHAVU_002G114200g	XP_007157976.1	29200	146	Carbamidomethylation; Deamidation (NQ)	1.30E+05	1.64E+05	0.00E+00	4.25E+05	1.27E+05	3.46E+03
hypothetical protein PHAVU_007G180800g	ESW16734.1	38845	144	Carbamidomethylation	2.46E+05	1.87E+04	0.00E+00	1.64E+05	8.21E+04	1.90E+04
hypothetical protein PHAVU_003G091300g	XP_007154111.1	31788	143	Carbamidomethylation	1.11E+05	2.01E+04	0.00E+00	9.10E+04	2.98E+04	7.58E+02
hypothetical protein PHAVU_007G057900g	ESW15259.1	26165	142		7.78E+04	4.76E+03	0.00E+00	1.31E+04	1.91E+05	0.00E+00
hypothetical protein PHAVU_006G093600g	XP_007147069.1	46270	140	Carbamidomethylation; Deamidation (NQ)	2.24E+04	2.10E+05	0.00E+00	2.02E+06	3.23E+05	5.90E+05
hypothetical protein PHAVU_002G286800g	ESW32025.1	39159	136		9.07E+03	1.71E+03	0.00E+00	5.45E+04	2.23E+03	0.00E+00
alcohol dehydrogenase 1	AGV54356.1	41355	136	Carbamidomethylation	2.63E+05	0.00E+00	0.00E+00	1.42E+05	6.06E+04	1.69E+04
triose-phosphate isomerase	CAI43251.1	27208	133	Carbamidomethylation	6.94E+04	6.91E+03	0.00E+00	6.15E+03	2.56E+04	0.00E+00

hypothetical protein PHAVU_003G154800g	XP_007154870.1	71141	132		1.31E+05	1.36E+04	0.00E+00	1.17E+04	4.94E+04	0.00E+00
hypothetical protein PHAVU_011G093100g	ESW04414.1	91143	131	Carbamidomethylation	3.67E+04	0.00E+00	0.00E+00	0.00E+00	6.50E+03	0.00E+00
hypothetical protein PHAVU_011G135600g	ESW04906.1	51376	129		1.30E+04	1.70E+03	0.00E+00	0.00E+00	2.14E+04	0.00E+00
hypothetical protein PHAVU_004G158800g	XP_007152777.1	26380	128		4.12E+04	4.97E+04	0.00E+00	4.31E+05	2.61E+05	0.00E+00
hypothetical protein PHAVU_009G172200g	XP_007137995.1	58562	126		1.36E+04	2.85E+04	0.00E+00	5.20E+04	0.00E+00	6.73E+02
alpha-amylase inhibitor like protein	BAA86927.1	28900	125		5.16E+05	8.43E+04	0.00E+00	1.10E+05	5.73E+05	2.05E+03
hypothetical protein PHAVU_003G121900g	ESW26464.1	39031	123		5.72E+04	5.25E+04	0.00E+00	3.03E+03	2.36E+04	5.44E+02
hypothetical protein PHAVU_008G281400g	XP_007142447.1	80056	122		4.16E+04	6.16E+03	0.00E+00	0.00E+00	7.47E+03	0.00E+00
hypothetical protein PHAVU_009G126800g	ESW09429.1	57515	121		3.64E+04	2.19E+04	0.00E+00	6.26E+03	2.48E+04	0.00E+00
hypothetical protein PHAVU_004G075100g	ESW23783.1	49331	116	Carbamidomethylation; Dioxidation (M)	5.94E+04	2.67E+04	0.00E+00	7.73E+04	5.54E+03	3.21E+03
hypothetical protein PHAVU_005G153300g	XP_007150435.1	35293	115	Carbamidomethylation	1.33E+04	0.00E+00	0.00E+00	1.31E+04	1.55E+04	0.00E+00
hypothetical protein PHAVU_003G051200g	XP_007153624.1	53217	113		2.25E+04	2.40E+03	0.00E+00	2.61E+03	5.35E+03	0.00E+00
hypothetical protein PHAVU_001G113800g	ESW33973.1	84083	112	Carbamidomethylation	8.60E+04	2.81E+03	0.00E+00	1.93E+03	1.12E+04	7.63E+03

hypothetical protein PHAVU_007G149400g	ESW16351.1	89807	110	Carbamidomethylation	1.48E+04	0.00E+00	0.00E+00	0.00E+00	1.10E+04	0.00E+00
hypothetical protein PHAVU_002G318400g	XP_007160397.1	63418	109		1.06E+04	0.00E+00	0.00E+00	4.70E+02	0.00E+00	0.00E+00
IAA-protein conjugate	AAG01035.2	35515	108		0.00E+00	7.56E+03	0.00E+00	4.06E+04	1.53E+04	0.00E+00
ADP-glucose pyrophosphorylase large subunit PvAGPL1	BAC66692.1	57835	108	Carbamidomethylation	2.51E+04	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.43E+03
nucleoredoxin 1-like protein	AGV54528.1	64765	107		6.44E+03	0.00E+00	0.00E+00	0.00E+00	4.66E+03	0.00E+00
hypothetical protein PHAVU_001G005200g	ESW32644.1	93840	107		3.20E+04	6.16E+03	0.00E+00	1.00E+04	5.38E+03	0.00E+00
hypothetical protein PHAVU_009G175800g	ESW10031.1	90868	107	Carbamidomethylation	1.80E+04	0.00E+00	0.00E+00	0.00E+00	1.10E+04	0.00E+00
hypothetical protein PHAVU_006G184200g	XP_007148147.1	26118	106		3.13E+03	7.43E+02	0.00E+00	0.00E+00	2.42E+04	0.00E+00
hypothetical protein PHAVU_008G189200g	ESW13355.1	38601	106		2.32E+04	3.57E+03	0.00E+00	0.00E+00	1.81E+04	0.00E+00
hypothetical protein PHAVU_007G034800g	ESW14989.1	59734	106		9.22E+03	6.47E+03	0.00E+00	1.54E+04	3.46E+03	0.00E+00
hypothetical protein PHAVU_010G136300g	XP_007135520.1	46660	105		7.20E+03	9.40E+02	0.00E+00	1.37E+04	0.00E+00	0.00E+00
hypothetical protein PHAVU_002G301400g	ESW32194.1	73421	104	Carbamidomethylation; Deamidation (NQ)	1.77E+04	5.76E+03	0.00E+00	0.00E+00	0.00E+00	0.00E+00

hypothetical protein PHAVU_011G164900g	ESW05257.1	27768	102	Carbamidomethylation	1.41E+04	0.00E+00	0.00E+00	6.71E+03	1.27E+04	0.00E+00
hypothetical protein PHAVU_005G051700g	ESW21214.1	63000	99		1.08E+04	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
hypothetical protein PHAVU_005G058100g	ESW21286.1	92590	98	Deamidation (NQ)	1.42E+04	0.00E+00	0.00E+00	2.73E+03	1.69E+03	0.00E+00
hypothetical protein PHAVU_007G162900g	XP_007144523.1	44653	97		1.83E+04	0.00E+00	0.00E+00	3.73E+03	0.00E+00	0.00E+00
hypothetical protein PHAVU_008G227900g	ESW13808.1	17782	96		4.12E+04	0.00E+00	0.00E+00	4.32E+03	5.89E+04	2.45E+03
cyclophilin	CAA52414.1	18160	96	Carbamidomethylation	1.67E+05	4.14E+03	0.00E+00	0.00E+00	1.26E+04	0.00E+00
hypothetical protein PHAVU_011G079400g	ESW04252.1	17850	96		6.19E+03	0.00E+00	0.00E+00	0.00E+00	3.18E+03	0.00E+00
hypothetical protein PHAVU_002G189300g	XP_007158877.1	28276	95		3.64E+04	1.42E+03	0.00E+00	1.01E+05	1.49E+04	0.00E+00
hypothetical protein PHAVU_005G035800g	ESW21036.1	19457	94		9.04E+03	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
hypothetical protein PHAVU_003G262900g	XP_007156155.1	28173	93		1.66E+04	3.05E+03	0.00E+00	0.00E+00	0.00E+00	0.00E+00
nucleoside diphosphate kinase 1	AGV54397.1	16552	93		5.01E+04	6.34E+03	0.00E+00	5.04E+03	1.23E+04	8.23E+03
hypothetical protein PHAVU_011G159000g	ESW05183.1	36024	92		9.59E+03	2.22E+03	0.00E+00	0.00E+00	0.00E+00	1.56E+04
formate dehydrogenase	ACZ74695.1	41265	90	Carbamidomethylation	4.98E+04	2.39E+04	0.00E+00	4.49E+04	9.54E+03	0.00E+00

hypothetical protein PHAVU_001G169000g	ESW34648.1	71349	90	Carbamidomethylation	4.49E+04	0.00E+00	0.00E+00	8.99E+03	2.34E+02	0.00E+00
ATP synthase subunit alpha mitochondrial	sp|P24459.1|ATPAM_PHAVU	55345	89		9.82E+03	5.17E+03	0.00E+00	1.61E+04	5.15E+03	0.00E+00
arcelin	CAD27954.1	30086	89		2.69E+05	6.01E+04	0.00E+00	6.85E+04	3.25E+05	1.16E+03
hypothetical protein PHAVU_003G069300g	ESW25836.1	29631	89		4.04E+03	1.92E+03	0.00E+00	6.95E+03	0.00E+00	4.05E+03
hypothetical protein PHAVU_005G050800g	ESW21204.1	42193	88		8.62E+03	3.26E+02	0.00E+00	0.00E+00	1.42E+04	0.00E+00
hypothetical protein PHAVU_011G167000g	ESW05279.1	35245	87	Carbamidomethylation	2.78E+05	0.00E+00	0.00E+00	0.00E+00	6.40E+04	4.94E+04
hypothetical protein PHAVU_008G013000g	ESW11234.1	71144	87	Carbamidomethylation	4.28E+04	0.00E+00	0.00E+00	0.00E+00	2.34E+02	0.00E+00
hypothetical protein PHAVU_011G061300g	XP_007132035.1	28137	87	Carbamidomethylation	4.47E+03	2.37E+03	0.00E+00	3.77E+03	1.03E+03	1.67E+04
RecName: Full=Arcelin-4; Flags: Precursor	sp|Q43629.1|ARC4_PHAVU	29451	86		3.75E+05	1.00E+05	0.00E+00	6.85E+04	5.17E+05	1.16E+03
hypothetical protein PHAVU_010G134700g	XP_007135504.1	95115	85		2.52E+05	4.99E+05	0.00E+00	4.62E+04	1.50E+04	0.00E+00
aspartic proteinase nepenthesin-1-like protein	AGV54394.1	46694	84		0.00E+00	2.43E+04	0.00E+00	0.00E+00	3.49E+03	7.69E+03
malate dehydrogenase	AGZ15381.1	35571	84	Carbamidomethylation	4.31E+04	1.59E+03	0.00E+00	0.00E+00	0.00E+00	0.00E+00

hypothetical protein PHAVU_005G115100g	ESW21967.1	107096	84		5.80E+03	0.00E+00	0.00E+00	0.00E+00	1.45E+03	0.00E+00
hypothetical protein PHAVU_006G126100g	XP_007147456.1	50306	84		6.02E+03	5.93E+02	0.00E+00	2.62E+03	6.96E+03	0.00E+00
glyceraldehyde-3-dehydrogenase C subunit	AGV54709.1	36619	83		4.64E+04	1.52E+03	0.00E+00	1.55E+04	2.30E+04	0.00E+00
hypothetical protein PHAVU_002G278400g	ESW31910.1	8178	82	Carbamidomethylation; Deamidation (NQ)	0.00E+00	1.50E+03	2.90E+03	1.13E+05	1.97E+03	2.04E+04
hypothetical protein PHAVU_010G075000g	ESW06770.1	10839	82		2.30E+03	0.00E+00	0.00E+00	0.00E+00	3.39E+03	0.00E+00
hypothetical protein PHAVU_002G309600g	ESW32289.1	71877	82		3.55E+03	0.00E+00	0.00E+00	0.00E+00	5.20E+03	0.00E+00
hypothetical protein PHAVU_011G008700g	XP_007131378.1	53212	82		0.00E+00	1.22E+05	0.00E+00	0.00E+00	2.02E+04	0.00E+00
hypothetical protein PHAVU_008G288400g	XP_007142529.1	41563	81	Carbamidomethylation	5.04E+02	0.00E+00	0.00E+00	0.00E+00	0.00E+00	2.12E+03
hypothetical protein PHAVU_008G081100g	ESW12057.1	46282	81	Carbamidomethylation	0.00E+00	7.62E+03	0.00E+00	0.00E+00	0.00E+00	0.00E+00
hypothetical protein PHAVU_007G276300g	XP_007145885.1	25290	81	Deamidation (NQ)	4.75E+04	8.15E+03	0.00E+00	2.12E+05	1.32E+04	0.00E+00
opper/zinc superoxide dismutase	AHA84142.1	15187	81	Carbamidomethylation	4.58E+04	3.46E+03	0.00E+00	3.33E+03	0.00E+00	0.00E+00
hypothetical protein PHAVU_011G183900g	ESW05492.1	18485	80		2.17E+04	0.00E+00	0.00E+00	2.22E+04	0.00E+00	0.00E+00

hypothetical protein PHAVU_010G135700g	XP_007135514.1	94696	79		5.87E+04	0.00E+00	0.00E+00	1.09E+04	0.00E+00	0.00E+00
hypothetical protein PHAVU_010G080300g	ESW06829.1	53602	78	Carbamidomethylation	0.00E+00	3.21E+03	0.00E+00	0.00E+00	0.00E+00	0.00E+00
hypothetical protein PHAVU_005G171400g	ESW22659.1	95630	78		7.57E+03	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
hypothetical protein PHAVU_006G170000g	ESW19968.1	27660	77	Carbamidomethylation	1.84E+04	0.00E+00	0.00E+00	1.98E+03	0.00E+00	0.00E+00
hypothetical protein PHAVU_001G114200g	XP_007161983.1	77300	76		1.46E+04	0.00E+00	0.00E+00	2.35E+03	0.00E+00	0.00E+00
hypothetical protein PHAVU_006G204400g	XP_007148388.1	58404	75		1.03E+04	5.45E+03	0.00E+00	2.31E+03	0.00E+00	3.30E+03
hypothetical protein PHAVU_008G228000g	ESW13809.1	17767	74	Deamidation (NQ)	2.32E+04	3.81E+03	0.00E+00	1.81E+04	2.81E+04	1.79E+03
hypothetical protein PHAVU_011G046100g	XP_007131845.1	23052	74		4.05E+03	2.96E+03	0.00E+00	2.50E+03	0.00E+00	0.00E+00
hypothetical protein PHAVU_008G056100g	XP_007139747.1	38760	74		5.86E+03	0.00E+00	0.00E+00	4.32E+03	2.57E+03	0.00E+00
hypothetical protein PHAVU_002G324300g	ESW32458.1	16757	74	Carbamidomethylation	4.64E+03	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
hypothetical protein PHAVU_003G113400g	ESW26364.1	26218	73	Carbamidomethylation	6.15E+03	0.00E+00	0.00E+00	0.00E+00	3.03E+03	4.40E+02
hypothetical protein PHAVU_005G044200g	ESW21130.1	8157	73	Carbamidomethylation	1.99E+03	0.00E+00	0.00E+00	0.00E+00	3.07E+03	0.00E+00
hypothetical protein PHAVU_002G280300g	ESW31936.1	62361	73	Carbamidomethylation	0.00E+00	1.61E+04	0.00E+00	6.68E+03	0.00E+00	0.00E+00

4-methyl-5(b-hydroxyethyl)-thiazole monophosphate biosynthesis protein	AHA84127.1	47093	72		0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.31E+03	0.00E+00
hypothetical protein PHAVU_007G145000g	XP_007144306.1	38963	72		7.68E+03	0.00E+00	0.00E+00	3.22E+03	2.31E+03	0.00E+00
alpha-1 4-glucan-protein synthase [UDP-forming]-like protein	AGV54358.1	41268	72	Carbamidomethylation	3.01E+03	0.00E+00	0.00E+00	1.08E+04	6.61E+03	2.98E+03
hypothetical protein PHAVU_007G109000g	XP_007143872.1	46094	69	Carbamidomethylation	1.01E+03	3.63E+03	0.00E+00	5.35E+02	1.09E+03	7.97E+03
hypothetical protein PHAVU_010G104800g	XP_007135145.1	16685	69		3.53E+04	2.92E+03	0.00E+00	0.00E+00	4.11E+04	0.00E+00
hypothetical protein PHAVU_008G096400g	ESW12243.1	12183	69		2.47E+03	6.30E+02	0.00E+00	0.00E+00	2.53E+03	0.00E+00
hypothetical protein PHAVU_005G048000g	XP_007149178.1	36660	69		7.96E+03	1.74E+03	0.00E+00	0.00E+00	7.04E+03	0.00E+00
hypothetical protein PHAVU_002G105000g	XP_007157870.1	98371	68		3.54E+03	0.00E+00	0.00E+00	6.62E+02	0.00E+00	0.00E+00
40S ribosomal protein S3-3-like protein	AGV54401.1	26724	68		9.48E+03	0.00E+00	0.00E+00	1.74E+04	0.00E+00	0.00E+00
hypothetical protein PHAVU_006G001600g	XP_007145963.1	99232	67		2.11E+03	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
hypothetical protein PHAVU_009G161600g	XP_007137859.1	83059	66		0.00E+00	2.05E+03	0.00E+00	0.00E+00	0.00E+00	0.00E+00
non-specific lipid transfer protein 1a precursor	ADC80502.1	11779	66	Carbamidomethylation	0.00E+00	0.00E+00	1.04E+03	0.00E+00	0.00E+00	0.00E+00

hypothetical protein PHAVU_011G166900g	ESW05278.1	36551	65	Carbamidomethylation	2.60E+04	1.96E+04	0.00E+00	0.00E+00	2.99E+04	3.84E+04
hypothetical protein PHAVU_005G073200g	XP_007149474.1	46480	65		0.00E+00	4.37E+03	0.00E+00	3.15E+03	0.00E+00	0.00E+00
40S ribosomal protein S5-like protein	AGV54378.1	23016	65		7.89E+02	0.00E+00	0.00E+00	0.00E+00	0.00E+00	7.76E+02
hypothetical protein PHAVU_007G041000g	XP_007143066.1	26172	65		0.00E+00	4.10E+03	0.00E+00	0.00E+00	0.00E+00	1.20E+03
hypothetical protein PHAVU_009G079500g	ESW08851.1	98347	64		3.53E+03	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
ubiquitin-like protein partial	AQR56342.1	14727	63		4.74E+03	1.96E+03	0.00E+00	7.96E+03	9.12E+03	0.00E+00
hypothetical protein PHAVU_011G072800g	ESW04175.1	55177	62		7.44E+03	0.00E+00	0.00E+00	1.65E+04	0.00E+00	0.00E+00
Pathogenesis-related protein 1	sp|P25985.2|PR1_PHAVU	16529	62	Deamidation (NQ)	4.34E+03	0.00E+00	0.00E+00	4.79E+03	1.50E+03	0.00E+00
hypothetical protein PHAVU_005G096300g	ESW21747.1	55354	62		2.60E+03	0.00E+00	0.00E+00	0.00E+00	1.57E+04	0.00E+00
hypothetical protein PHAVU_007G055800g	ESW15232.1	97762	61	Dioxidation (M)	1.50E+04	0.00E+00	0.00E+00	1.09E+04	0.00E+00	0.00E+00
hypothetical protein PHAVU_005G056100g	ESW21264.1	9608	60	Deamidation (NQ)	0.00E+00	0.00E+00	0.00E+00	0.00E+00	6.40E+03	0.00E+00
hypothetical protein PHAVU_003G156100g	ESW26877.1	59534	60	Carbamidomethylation	3.27E+03	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
14-3-3 protein	AGV54448.1	29160	60		1.09E+03	0.00E+00	0.00E+00	3.18E+03	4.17E+03	0.00E+00

elongation factor 1-beta	AGV54725.1	24287	60		1.63E+03	0.00E+00	0.00E+00	0.00E+00	1.15E+04	0.00E+00
hypothetical protein PHAVU_002G294200g	XP_007160120.1	95203	59		1.72E+03	0.00E+00	0.00E+00	1.89E+03	2.70E+02	0.00E+00
hypothetical protein PHAVU_009G139500g	XP_007137596.1	63211	59		1.16E+04	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
GNL-2 alpha subunit=lectin {N-terminal}	AAB36314.1	2407	59		5.12E+05	4.74E+05	0.00E+00	1.57E+05	9.06E+05	8.68E+04
hypothetical protein PHAVU_007G250300g	XP_007145577.1	68993	59		0.00E+00	5.15E+03	0.00E+00	0.00E+00	0.00E+00	0.00E+00
hypothetical protein PHAVU_003G062900g	ESW25754.1	163483	58		8.56E+03	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
hypothetical protein PHAVU_002G143600g	XP_007158331.1	88224	58		1.54E+04	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
hypothetical protein PHAVU_010G125100g	XP_007135385.1	33888	57		9.58E+03	0.00E+00	0.00E+00	7.97E+03	0.00E+00	0.00E+00
hypothetical protein PHAVU_001G266600g	XP_007163814.1	51672	57	Carbamidomethylation	3.71E+03	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
hypothetical protein PHAVU_010G082900g	XP_007134868.1	17151	57	Carbamidomethylation	4.15E+03	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
hypothetical protein PHAVU_003G268700g	XP_007156223.1	40561	55		2.37E+03	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
hypothetical protein PHAVU_009G194100g	XP_007138265.1	21459	54		4.11E+03	0.00E+00	0.00E+00	1.90E+04	0.00E+00	2.91E+03

aquaporin PIP2	AGV54658.1	30841	53		0.00E+00	6.56E+02	0.00E+00	0.00E+00	0.00E+00	1.09E+03
hypothetical protein PHAVU_009G025100g	XP_007136182.1	18700	53		1.57E+03	1.82E+03	0.00E+00	2.48E+03	1.32E+03	0.00E+00
transaldolase-like protein	AGV54320.1	48418	52		0.00E+00	0.00E+00	0.00E+00	0.00E+00	5.74E+02	0.00E+00
DNA-binding protein GBP16	AHA84129.1	43458	52		2.14E+03	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
Ferritin chloroplastic	sp|P25699.1|FRI_PHAVU	28304	52		2.12E+03	2.50E+03	0.00E+00	0.00E+00	3.78E+03	0.00E+00
hypothetical protein PHAVU_003G208200g	XP_007155515.1	44548	51		0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.21E+03	0.00E+00
hypothetical protein PHAVU_005G032000g	ESW20993.1	10653	51		0.00E+00	0.00E+00	0.00E+00	0.00E+00	6.36E+03	0.00E+00
hypothetical protein PHAVU_007G135100g	ESW16174.1	61259	51		0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
hypothetical protein PHAVU_008G156300g	ESW12963.1	11354	50		1.27E+04	1.56E+03	0.00E+00	9.94E+03	0.00E+00	0.00E+00
hypothetical protein PHAVU_003G060400g	ESW25726.1	28104	50	Carbamidomethylation	4.67E+03	0.00E+00	0.00E+00	0.00E+00	2.85E+03	0.00E+00
hypothetical protein PHAVU_009G177000g	XP_007138058.1	13535	50		1.37E+04	0.00E+00	0.00E+00	9.94E+03	0.00E+00	0.00E+00
hypothetical protein PHAVU_009G255100g	XP_007138986.1	18815	49		0.00E+00	0.00E+00	0.00E+00	2.09E+03	0.00E+00	0.00E+00
hypothetical protein PHAVU_004G073800g	ESW23769.1	27237	49		3.86E+03	0.00E+00	0.00E+00	0.00E+00	4.95E+03	0.00E+00

hypothetical protein PHAVU_009G141600g	XP_007137617.1	26575	49		2.17E+03	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
hypothetical protein PHAVU_004G000800g	ESW22861.1	55255	48	Carbamidomethylation	2.64E+03	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
hypothetical protein PHAVU_005G174800g	XP_007150713.1	14644	48		7.67E+03	9.15E+02	0.00E+00	0.00E+00	5.96E+03	0.00E+00
hypothetical protein PHAVU_005G161300g	XP_007150544.1	16839	48		1.56E+03	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
hypothetical protein PHAVU_005G059300g	ESW21300.1	25548	48		8.06E+02	0.00E+00	0.00E+00	0.00E+00	9.39E+02	0.00E+00
14-3-3 protein	AGV54291.1	29314	48		0.00E+00	0.00E+00	0.00E+00	2.97E+03	1.79E+03	0.00E+00
hypothetical protein PHAVU_009G106800g	ESW09178.1	44948	47	Carbamidomethylation	2.56E+03	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
hypothetical protein PHAVU_003G2836000g partial	XP_007156409.1	22866	47		4.32E+02	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
hypothetical protein PHAVU_011G033900g	ESW03686.1	18901	46		3.10E+03	0.00E+00	0.00E+00	2.90E+03	0.00E+00	0.00E+00
Bowman-Birk type proteinase inhibitor 2	sp|P01060.3|IBB2_PHAVU	11637	46	Carbamidomethylation	4.43E+03	9.64E+02	0.00E+00	1.69E+04	7.65E+03	1.94E+04
albumin-1B	ADR30069.1	13771	46	Carbamidomethylation	0.00E+00	0.00E+00	4.62E+05	0.00E+00	0.00E+00	0.00E+00
hypothetical protein PHAVU_007G206700g	XP_007145063.1	83277	45		3.22E+02	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00

hypothetical protein PHAVU_007G230300g	XP_007145333.1	46915	45	Carbamidomethylation	3.19E+03	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
60S acidic ribosomal protein P3-like protein	AGV54398.1	11890	45		1.95E+03	6.43E+02	0.00E+00	0.00E+00	1.91E+03	0.00E+00
hypothetical protein PHAVU_003G1967000g partial	XP_007155385.1	50769	44		3.33E+03	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
GTP-binding nuclear protein Ran-3	AGV54506.1	25106	44	Carbamidomethylation	3.21E+03	0.00E+00	0.00E+00	0.00E+00	1.70E+03	0.00E+00
cytosolic glutathione reductase	ABF29524.1	54762	44		4.03E+02	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
hypothetical protein PHAVU_001G003200g	XP_007160627.1	34546	43		9.39E+02	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
LEA-18	AAF81194.1	8777	42		0.00E+00	0.00E+00	0.00E+00	0.00E+00	2.54E+04	0.00E+00
hypothetical protein PHAVU_L002900g	XP_007163953.1	17798	42		0.00E+00	0.00E+00	0.00E+00	0.00E+00	2.09E+03	0.00E+00
hypothetical protein PHAVU_011G066700g	ESW04093.1	23664	42		4.04E+03	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
hypothetical protein PHAVU_007G145200g	XP_007144308.1	15060	42		0.00E+00	0.00E+00	0.00E+00	4.77E+04	0.00E+00	0.00E+00
hypothetical protein PHAVU_003G209800g	ESW27526.1	52552	42		1.26E+03	0.00E+00	0.00E+00	4.56E+02	5.38E+02	0.00E+00
hypothetical protein PHAVU_007G146800g	XP_007144328.1	13538	41		1.29E+03	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00

RNA-binding region RNP-1	AHA84227.1	29654	41		9.94E+02	0.00E+00	0.00E+00	0.00E+00	6.01E+02	0.00E+00
cysteine proteinase precursor	CAB17076.1	50253	41		0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.11E+03	6.62E+02
hypothetical protein PHAVU_002G029900g	XP_007156938.1	14021	41		1.63E+03	0.00E+00	0.00E+00	0.00E+00	3.57E+03	0.00E+00
hypothetical protein PHAVU_001G157800g	ESW34501.1	50796	41		0.00E+00	0.00E+00	0.00E+00	7.55E+02	0.00E+00	0.00E+00
hypothetical protein PHAVU_007G245700g	ESW17518.1	13724	40		0.00E+00	0.00E+00	0.00E+00	0.00E+00	4.62E+03	0.00E+00
hypothetical protein PHAVU_008G269300g	ESW14298.1	50397	40		0.00E+00	0.00E+00	0.00E+00	2.62E+03	0.00E+00	0.00E+00
hypothetical protein PHAVU_009G224900g	ESW10624.1	26325	40		0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.77E+03	0.00E+00
hypothetical protein PHAVU_010G008900g	ESW05975.1	18056	38		2.27E+03	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
hypothetical protein PHAVU_001G186100g	XP_007162849.1	47661	38		0.00E+00	0.00E+00	0.00E+00	0.00E+00	5.41E+03	0.00E+00
hypothetical protein PHAVU_009G241700g	XP_007138839.1	15658	38		1.63E+03	6.44E+02	0.00E+00	0.00E+00	0.00E+00	3.73E+02
			38							
putative calcium binding protein partial	AAZ23153.1	12209	38		0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
steroid binding protein	AGV54313.1	25342	38		1.12E+03	6.40E+02	0.00E+00	0.00E+00	0.00E+00	0.00E+00

hypothetical protein PHAVU_002G288700g	ESW32047.1	25613	37		1.17E+04	4.25E+03	0.00E+00	0.00E+00	0.00E+00	0.00E+00
hypothetical protein PHAVU_001G064000g	ESW33374.1	41037	37		5.75E+02	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
hypothetical protein PHAVU_007G100200g partial	XP_007143771.1	16737	37		0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
60S acidic ribosomal protein P0	AGV54296.1	34186	37		4.16E+03	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
hypothetical protein PHAVU_009G254300g	ESW10969.1	83596	37		0.00E+00	0.00E+00	0.00E+00	1.85E+03	0.00E+00	0.00E+00
hypothetical protein PHAVU_003G209400g	XP_007155527.1	26610	36		7.83E+02	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
hypothetical protein PHAVU_003G035400g	ESW25434.1	102562	36		0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
hypothetical protein PHAVU_005G178300g	ESW22751.1	44625	35.9		4.53E+03	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
-10lgP: Score from Peaks Xpro software										

Supplementary Table S2.2. Compositions, neutral mass, and retention time (minutes) of all oligosaccharides confirmed by tandem MS/MS in at least one of the bean subcritical water extractions.

Hex_dHex_Pent_HexA	Mass	Retention time (min)	Hex	dHex	Pent	HexA
0_1_1_0	312.106	6.6	0	1	1	0
1_0_1_0	326.121	4.8	1	0	1	0
0_0_3_0	414.137	12.7	0	0	3	0
0_0_3_0	414.137	13.7	0	0	3	0
0_0_3_0	414.137	16.7	0	0	3	0
0_0_3_0	414.137	17.6	0	0	3	0
1_2_0_0	472.179	12.0	1	2	0	0
2_0_1_0	474.159	3.0	2	0	1	0
2_0_1_0	474.159	4.5	2	0	1	0
2_0_1_0	474.159	12.3	2	0	1	0
2_0_1_0	474.159	15.7	2	0	1	0
2_0_1_0	474.159	17.5	2	0	1	0
2_0_1_0	474.159	18.5	2	0	1	0
2_0_1_0	474.159	19.5	2	0	1	0
2_1_0_0	488.174	7.0	2	1	0	0

2_1_0_0	488.174	14.4	2	1	0	0
3_0_0_0	504.169	11.6	3	0	0	0
3_0_0_0	504.169	13.8	3	0	0	0
3_0_0_0	504.170	7.2	3	0	0	0
3_0_0_0	504.170	9.0	3	0	0	0
3_0_0_0	504.170	12.5	3	0	0	0
3_0_0_0	504.170	13.0	3	0	0	0
3_0_0_0	504.170	14.5	3	0	0	0
3_0_0_0	504.170	16.3	3	0	0	0
3_0_0_0	504.170	17.5	3	0	0	0
3_0_0_0	504.170	18.5	3	0	0	0
3_0_0_0	504.170	19.5	3	0	0	0
3_0_0_0	504.170	20.0	3	0	0	0
3_0_0_0	504.170	21.5	3	0	0	0
3_0_0_0	504.170	22.5	3	0	0	0
2_0_0_1	518.148	16.0	2	0	0	1
1_1_0_1	520.164	15.0	1	1	0	1
1_1_0_1	520.164	21.0	1	1	0	1

0_0_4_0	546.180	11.0	0	0	4	0
0_0_4_0	546.180	12.0	0	0	4	0
0_0_4_0	546.180	14.7	0	0	4	0
0_0_4_0	546.180	16.5	0	0	4	0
0_0_4_0	546.180	17.5	0	0	4	0
0_0_4_0	546.180	18.2	0	0	4	0
0_0_4_0	546.180	19.5	0	0	4	0
0_0_4_0	546.180	21.0	0	0	4	0
0_0_4_0	546.180	22.0	0	0	4	0
0_0_4_0	546.180	23.0	0	0	4	0
0_0_4_0	546.180	24.0	0	0	4	0
0_0_4_0	546.180	25.5	0	0	4	0
0_0_4_0	546.180	26.0	0	0	4	0
0_0_4_0	546.180	28.0	0	0	4	0
0_2_2_0	574.211	6.5	0	2	2	0
1_0_3_0	576.190	4.0	1	0	3	0
1_0_3_0	576.190	5.8	1	0	3	0
1_1_2_0	590.206	4.0	1	1	2	0

2_0_2_0	606.201	3.3	2	0	2	0
2_0_2_0	606.201	3.7	2	0	2	0
2_0_2_0	606.201	4.1	2	0	2	0
2_0_2_0	606.201	4.5	2	0	2	0
2_0_2_0	606.201	4.7	2	0	2	0
2_0_2_0	606.201	5.7	2	0	2	0
2_0_2_0	606.201	15.3	2	0	2	0
2_0_2_0	606.201	17.0	2	0	2	0
2_0_2_0	606.201	18.3	2	0	2	0
2_0_2_0	606.201	21.2	2	0	2	0
2_0_2_0	606.201	21.9	2	0	2	0
2_0_2_0	606.201	26.7	2	0	2	0
1_1_0_0	634.232	12.0	1	1	0	0
3_0_1_0	636.211	3.4	3	0	1	0
3_0_1_0	636.211	3.4	3	0	1	0
3_0_1_0	636.211	4.3	3	0	1	0
3_0_1_0	636.211	9.0	3	0	1	0
3_0_1_0	636.211	10.2	3	0	1	0

3_0_1_0	636.211	11.2	3	0	1	0
3_0_1_0	636.211	11.7	3	0	1	0
3_0_1_0	636.211	11.8	3	0	1	0
3_0_1_0	636.211	12.1	3	0	1	0
3_0_1_0	636.211	12.5	3	0	1	0
3_0_1_0	636.211	13.0	3	0	1	0
3_0_1_0	636.211	14.5	3	0	1	0
3_0_1_0	636.211	15.7	3	0	1	0
3_0_1_0	636.211	16.2	3	0	1	0
3_0_1_0	636.211	17.5	3	0	1	0
3_0_1_0	636.211	18.0	3	0	1	0
3_0_1_0	636.211	18.8	3	0	1	0
3_0_1_0	636.211	20.1	3	0	1	0
3_0_1_0	636.211	21.0	3	0	1	0
3_0_1_0	636.211	23.3	3	0	1	0
3_0_1_0	636.211	24.0	3	0	1	0
3_0_1_0	636.211	24.6	3	0	1	0
3_0_1_0	636.211	25.0	3	0	1	0

3_0_1_0	636.211	25.6	3	0	1	0
1_2_0_1	648.211	5.9	1	2	0	1
1_2_0_1	648.211	13.8	1	2	0	1
1_2_0_1	648.211	14.6	1	2	0	1
1_2_0_1	648.211	15.0	1	2	0	1
1_2_0_1	648.211	15.5	1	2	0	1
1_2_0_1	648.211	21.5	1	2	0	1
1_2_0_1	648.211	23.8	1	2	0	1
2_0_1_1	650.191	19.2	2	0	1	1
3_1_0_0	650.227	6.7	3	1	0	0
3_1_0_0	650.227	10.8	3	1	0	0
3_1_0_0	650.227	12.0	3	1	0	0
3_1_0_0	650.227	14.5	3	1	0	0
3_1_0_0	650.227	15.7	3	1	0	0
3_0_1_0 +H2O	654.222	3.4	3	0	1	0
2_1_0_1	664.206	14.0	2	1	0	1
4_0_0_0	666.222	3.0	4	0	0	0
4_0_0_0	666.222	4.0	4	0	0	0

4_0_0_0	666.222	8.9	4	0	0	0
4_0_0_0	666.222	10.0	4	0	0	0
4_0_0_0	666.222	14.7	4	0	0	0
4_0_0_0	666.222	16.0	4	0	0	0
4_0_0_0	666.222	17.0	4	0	0	0
4_0_0_0	666.222	17.7	4	0	0	0
4_0_0_0	666.222	19.1	4	0	0	0
4_0_0_0	666.222	20.2	4	0	0	0
4_0_0_0	666.222	23.0	4	0	0	0
4_0_0_0	666.222	24.6	4	0	0	0
4_0_0_0	666.222	26.1	4	0	0	0
4_0_0_0	666.222	26.7	4	0	0	0
4_0_0_0	666.222	27.5	4	0	0	0
0_0_5_0	678.222	14.4	0	0	5	0
0_0_5_0	678.222	15.0	0	0	5	0
0_0_5_0	678.222	16.5	0	0	5	0
0_0_5_0	678.222	20.1	0	0	5	0
0_0_5_0	678.222	23.0	0	0	5	0

0_0_5_0	678.222	24.8	0	0	5	0
0_0_5_0	678.222	25.3	0	0	5	0
0_0_5_0	678.222	26.0	0	0	5	0
0_0_5_0	678.222	28.1	0	0	5	0
0_0_5_0	678.222	30.1	0	0	5	0
0_0_5_0	678.222	11.6	0	0	5	0
0_0_5_0	678.222	17.5	0	0	5	0
0_0_5_0	678.222	21.6	0	0	5	0
3_0_0_1	680.201	17.9	3	0	0	1
3_0_0_1	680.201	18.7	3	0	0	1
3_0_0_1	680.201	19.2	3	0	0	1
1_0_4_0	708.233	5.5	1	0	4	0
1_0_4_0	708.233	15.5	1	0	4	0
1_0_4_0	708.233	21.0	1	0	4	0
0_0_4_1	722.212	25.2	0	0	4	1
2_0_3_0	738.243	21.6	2	0	3	0
0_2_2_1	750.243	5.0	0	2	2	1
3_0_2_0	768.254	10.2	3	0	2	0

3_0_2_0	768.254	11.5	3	0	2	0
3_0_2_0	768.254	19.1	3	0	2	0
3_0_2_0	768.254	20.4	3	0	2	0
3_0_2_0	768.254	21.8	3	0	2	0
3_0_2_0	768.254	22.5	3	0	2	0
3_0_2_0	768.254	24.6	3	0	2	0
3_0_2_0	768.254	25.0	3	0	2	0
3_0_2_0	768.254	26.2	3	0	2	0
0_2_2_1 +H2O	768.254	5.0	0	2	2	1
1_2_1_1	780.254	24.6	1	2	1	1
2_0_2_1	782.233	28.5	2	0	2	1
2_0_2_1	782.233	30.6	2	0	2	1
3_1_1_0	782.269	6.9	3	1	1	0
4_0_1_0	798.264	12.1	4	0	1	0
4_0_1_0	798.264	15.4	4	0	1	0
4_0_1_0	798.264	16.5	4	0	1	0
4_0_1_0	798.264	17.8	4	0	1	0
4_0_1_0	798.264	19.0	4	0	1	0

4_0_1_0	798.264	19.9	4	0	1	0
4_0_1_0	798.264	27.3	4	0	1	0
4_0_1_0	798.264	28.4	4	0	1	0
4_0_1_0	798.264	29.0	4	0	1	0
3_1_1_0 +H2O	800.280	6.9	3	1	1	0
0_3_0_2	808.249	21.2	0	3	0	2
2_2_0_1	810.264	16.5	2	2	0	1
0_0_6_0	810.264	11.1	0	0	6	0
0_0_6_0	810.264	13.1	0	0	6	0
0_0_6_0	810.264	13.4	0	0	6	0
0_0_6_0	810.264	14.0	0	0	6	0
0_0_6_0	810.264	14.5	0	0	6	0
0_0_6_0	810.264	15.7	0	0	6	0
0_0_6_0	810.264	16.6	0	0	6	0
0_0_6_0	810.264	17.5	0	0	6	0
0_0_6_0	810.264	23.0	0	0	6	0
0_0_6_0	810.264	24.7	0	0	6	0
0_0_6_0	810.264	25.9	0	0	6	0

0_0_6_0	810.264	27.3	0	0	6	0
0_0_6_0	810.264	28.0	0	0	6	0
0_0_6_0	810.264	29.0	0	0	6	0
0_0_6_0	810.264	29.8	0	0	6	0
0_0_6_0	810.264	30.3	0	0	6	0
4_1_0_0	812.280	16.4	4	1	0	0
3_1_0_1	826.259	19.3	3	1	0	1
5_0_0_0	828.275	17.6	5	0	0	0
5_0_0_0	828.275	21.6	5	0	0	0
5_0_0_0	828.275	9.9	5	0	0	0
5_0_0_0	828.275	11.0	5	0	0	0
5_0_0_0	828.275	11.8	5	0	0	0
5_0_0_0	828.275	16.6	5	0	0	0
5_0_0_0	828.275	17.4	5	0	0	0
5_0_0_0	828.275	18.1	5	0	0	0
5_0_0_0	828.275	18.9	5	0	0	0
5_0_0_0	828.275	19.5	5	0	0	0
5_0_0_0	828.275	23.0	5	0	0	0

5_0_0_0	828.275	24.8	5	0	0	0
5_0_0_0	828.275	29.5	5	0	0	0
5_0_0_0	828.275	30.0	5	0	0	0
5_0_0_0	828.275	31.1	5	0	0	0
4_1_0_0 +H2O	830.290	12.3	4	1	0	0
3_1_0_1 +H2O	844.270	19.6	3	1	0	1
2_0_4_0	870.285	15.3	2	0	4	0
2_0_4_0	870.285	16.1	2	0	4	0
3_0_3_0	900.296	10.1	3	0	3	0
3_0_3_0	900.296	10.6	3	0	3	0
3_0_3_0	900.296	11.5	3	0	3	0
3_0_3_0	900.296	12.9	3	0	3	0
3_0_3_0	900.296	14.6	3	0	3	0
3_0_3_0	900.296	22.6	3	0	3	0
3_0_3_0	900.296	23.1	3	0	3	0
3_0_3_0 +H2O	918.307	12.7	3	0	3	0
3_0_3_0 +H2O	918.307	13.5	3	0	3	0
4_0_2_0	930.306	10.1	4	0	2	0

4_0_2_0	930.306	10.5	4	0	2	0
4_0_2_0	930.306	17.0	4	0	2	0
4_0_2_0	930.306	25.6	4	0	2	0
4_0_2_0	930.306	29.4	4	0	2	0
0_0_7_0	942.307	15.7	0	0	7	0
0_0_7_0	942.307	22.2	0	0	7	0
0_0_7_0	942.307	23.6	0	0	7	0
0_0_7_0	942.307	24.2	0	0	7	0
0_0_7_0	942.307	24.8	0	0	7	0
0_0_7_0	942.307	25.6	0	0	7	0
0_0_7_0	942.307	26.0	0	0	7	0
0_0_7_0	942.307	26.5	0	0	7	0
0_0_7_0	942.307	29.0	0	0	7	0
0_0_7_0	942.307	29.8	0	0	7	0
0_0_7_0	942.307	30.0	0	0	7	0
5_0_1_0	960.317	12.2	5	0	1	0
5_0_1_0	960.317	20.2	5	0	1	0
5_0_1_0	960.317	20.9	5	0	1	0

5_0_1_0	960.317	22.1	5	0	1	0
0_2_5_0	970.338	26.3	0	2	5	0
3_2_0_1	972.317	10.6	3	2	0	1
3_2_0_1	972.317	11.9	3	2	0	1
3_2_0_1	972.317	13.1	3	2	0	1
1_1_5_0	986.333	21.7	1	1	5	0
6_0_0_0	990.327	6.7	6	0	0	0
6_0_0_0	990.327	8.6	6	0	0	0
6_0_0_0	990.327	9.5	6	0	0	0
6_0_0_0	990.327	10.1	6	0	0	0
6_0_0_0	990.327	10.5	6	0	0	0
6_0_0_0	990.327	11.4	6	0	0	0
6_0_0_0	990.327	11.9	6	0	0	0
6_0_0_0	990.327	12.1	6	0	0	0
6_0_0_0	990.327	13.1	6	0	0	0
6_0_0_0	990.327	13.8	6	0	0	0
6_0_0_0	990.327	14.3	6	0	0	0
6_0_0_0	990.327	15.3	6	0	0	0

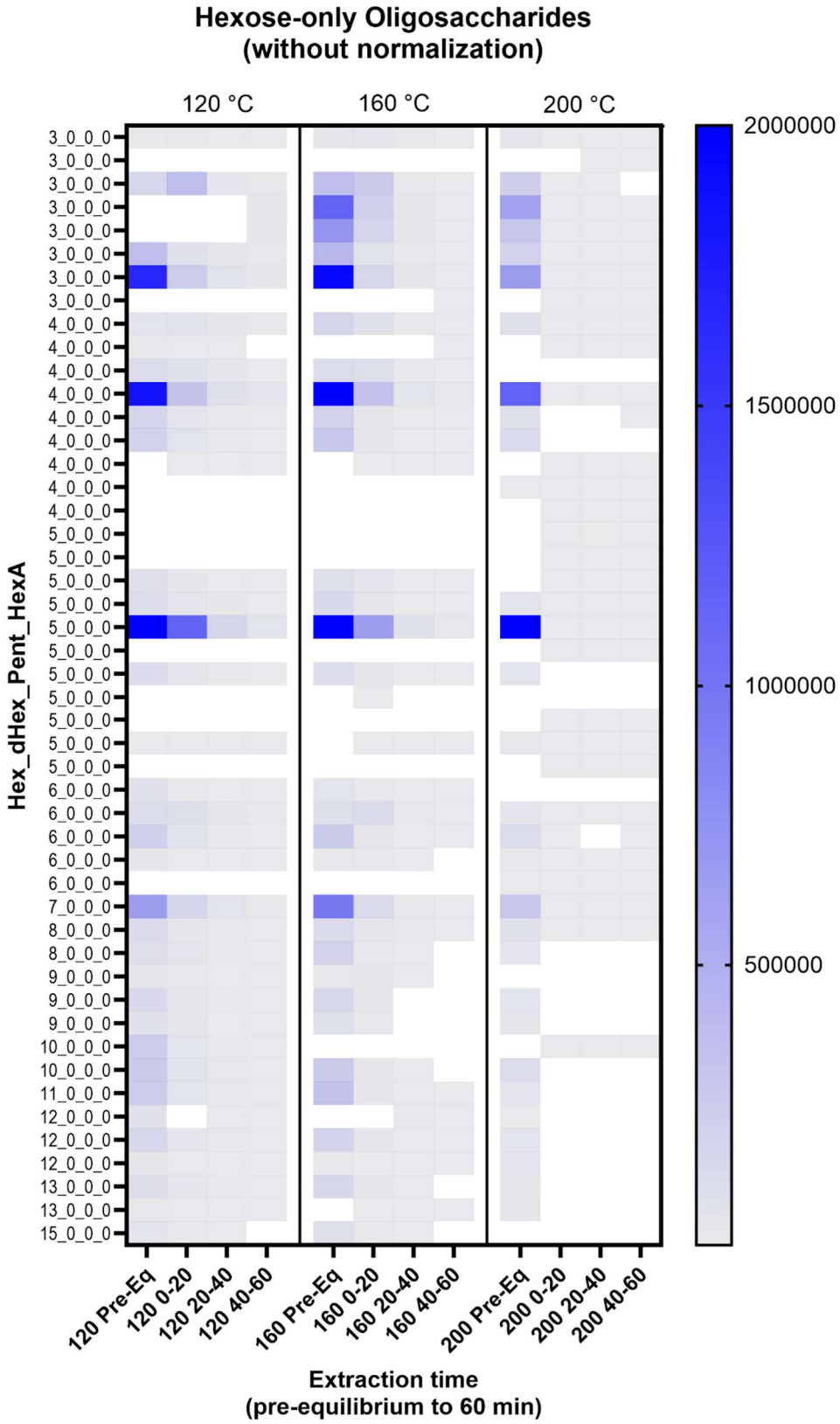
6_0_0_0	990.327	18.3	6	0	0	0
6_0_0_0	990.327	19.2	6	0	0	0
6_0_0_0	990.327	20.2	6	0	0	0
6_0_0_0	990.327	21.0	6	0	0	0
6_0_0_0	990.327	22.6	6	0	0	0
6_0_0_0	990.327	29.0	6	0	0	0
6_0_0_0	990.327	30.6	6	0	0	0
6_0_0_0	990.327	33.4	6	0	0	0
3_0_4_0	1032.338	21.1	3	0	4	0
3_0_4_0	1032.338	25.6	3	0	4	0
4_0_3_0	1062.349	14.2	4	0	3	0
4_0_3_0	1062.349	15.3	4	0	3	0
4_0_3_0	1062.349	15.9	4	0	3	0
4_0_3_0	1062.349	17.0	4	0	3	0
0_0_8_0	1074.349	21.0	0	0	8	0
0_0_8_0	1074.349	22.0	0	0	8	0
0_0_8_0	1074.349	22.5	0	0	8	0
0_0_8_0	1074.349	23.3	0	0	8	0

0_0_8_0	1074.349	25.5	0	0	8	0
0_0_8_0	1074.349	26.3	0	0	8	0
0_0_8_0	1074.349	28.0	0	0	8	0
0_0_8_0	1074.349	29.4	0	0	8	0
0_0_8_0	1074.349	31.3	0	0	8	0
0_0_8_0	1074.349	32.6	0	0	8	0
4_0_3_0 +H2O	1080.359	15.8	4	0	3	0
0_0_8_0	1092.360	21.2	0	0	8	0
0_0_8_0	1092.360	22.2	0	0	8	0
0_2_6_0	1102.380	28.0	0	2	6	0
7_0_0_0	1152.380	19.9	7	0	0	0
7_0_0_0	1152.380	20.5	7	0	0	0
7_0_0_0	1152.380	21.2	7	0	0	0
7_0_0_0	1152.380	22.2	7	0	0	0
7_0_0_0	1152.380	22.9	7	0	0	0
7_0_0_0	1152.380	23.5	7	0	0	0
0_0_9_0	1206.391	14.9	0	0	9	0
0_0_9_0	1206.391	26.1	0	0	9	0

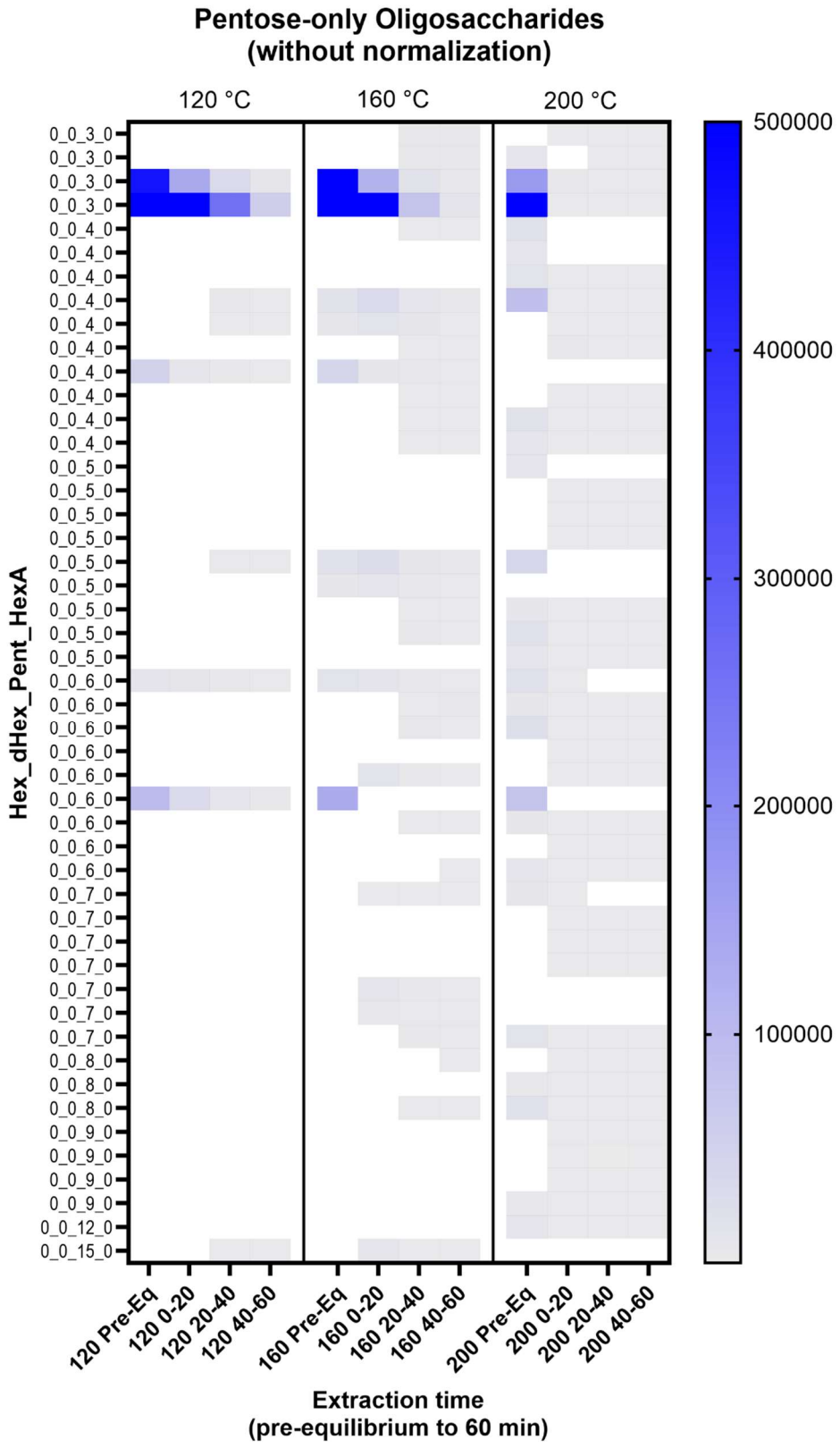
0_0_9_0	1206.391	28.4	0	0	9	0
0_0_9_0	1206.391	30.5	0	0	9	0
1_3_1_3	1278.376	26.1	1	3	1	3
8_0_0_0	1314.433	16.3	8	0	0	0
8_0_0_0	1314.433	17.2	8	0	0	0
8_0_0_0	1314.433	20.4	8	0	0	0
8_0_0_0	1314.433	21.2	8	0	0	0
8_0_0_0	1314.433	22.6	8	0	0	0
8_0_0_0	1314.433	23.1	8	0	0	0
8_0_0_0	1314.433	24.0	8	0	0	0
8_0_0_0 +H2O	1332.444	15.8	8	0	0	0
8_0_0_0 +H2O	1332.444	17.8	8	0	0	0
0_0_10_0 -H2O	1338.433	27.0	0	0	10	0
9_0_0_0	1476.486	10.7	9	0	0	0
9_0_0_0	1476.486	21.9	9	0	0	0
9_0_0_0	1476.486	22.6	9	0	0	0
9_0_0_0	1476.486	23.3	9	0	0	0
9_0_0_0	1476.486	23.8	9	0	0	0

9_0_0_0	1476.486	24.4	9	0	0	0
9_0_0_0 +H2O	1494.496	9.6	9	0	0	0
0_0_12_0	1602.518	27.7	0	0	12	0
10_0_0_0	1638.539	23.2	10	0	0	0
10_0_0_0	1638.539	24.4	10	0	0	0
0_0_13_0	1734.560	28.7	0	0	13	0
11_0_0_0	1800.591	24.7	11	0	0	0
11_0_0_0	1800.591	26.0	11	0	0	0
12_0_0_0	1962.644	18.7	12	0	0	0
12_0_0_0	1962.644	20.4	12	0	0	0
12_0_0_0	1962.644	24.5	12	0	0	0
12_0_0_0	1962.644	25.4	12	0	0	0
12_0_0_0	1962.644	26.8	12	0	0	0
0_0_15_0	1998.644	29.5	0	0	15	0
13_0_0_0	2124.697	25.0	13	0	0	0
13_0_0_0	2124.697	25.7	13	0	0	0
15_0_0_0	2448.803	26.6	15	0	0	0

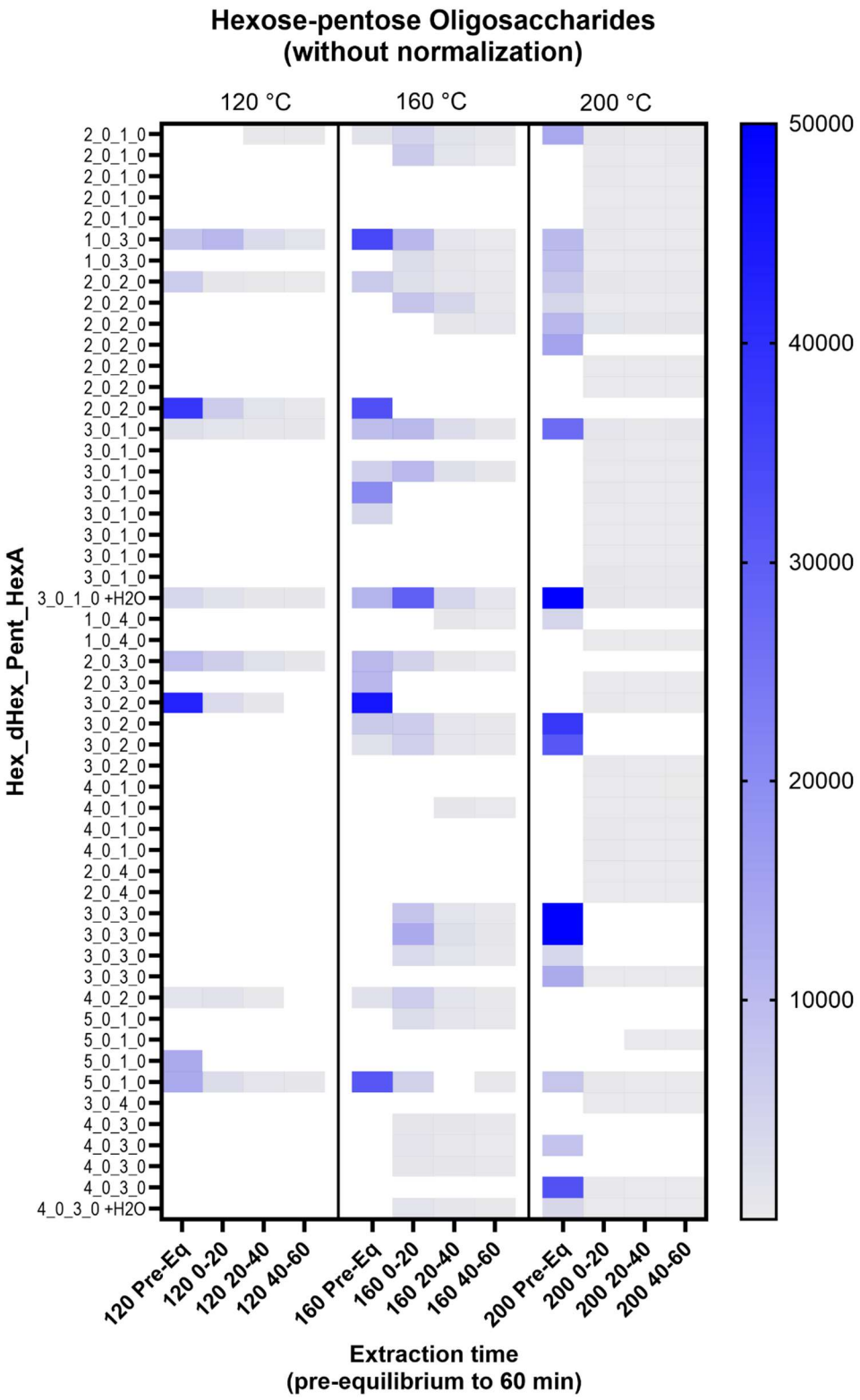
A)


Hexose-only Oligosaccharides
(without normalization)
120 °C
160 °C
200 °C
Hex_dHex_Pent_HexA
3_0_0_0
4_0_0_0
5_0_0_0
6_0_0_0
7_0_0_0
8_0_0_0
9_0_0_0
10_0_0_0
11_0_0_0
12_0_0_0
13_0_0_0
15_0_0_0
2000000
1500000
1000000
500000
120 Pre-Eq
120 0-20
120 20-40
120 40-60
160 Pre-Eq
160 0-20
160 20-40
160 40-60
200 Pre-Eq
200 0-20
200 20-40
200 40-60
Extraction time
(pre-equilibrium to 60 min)

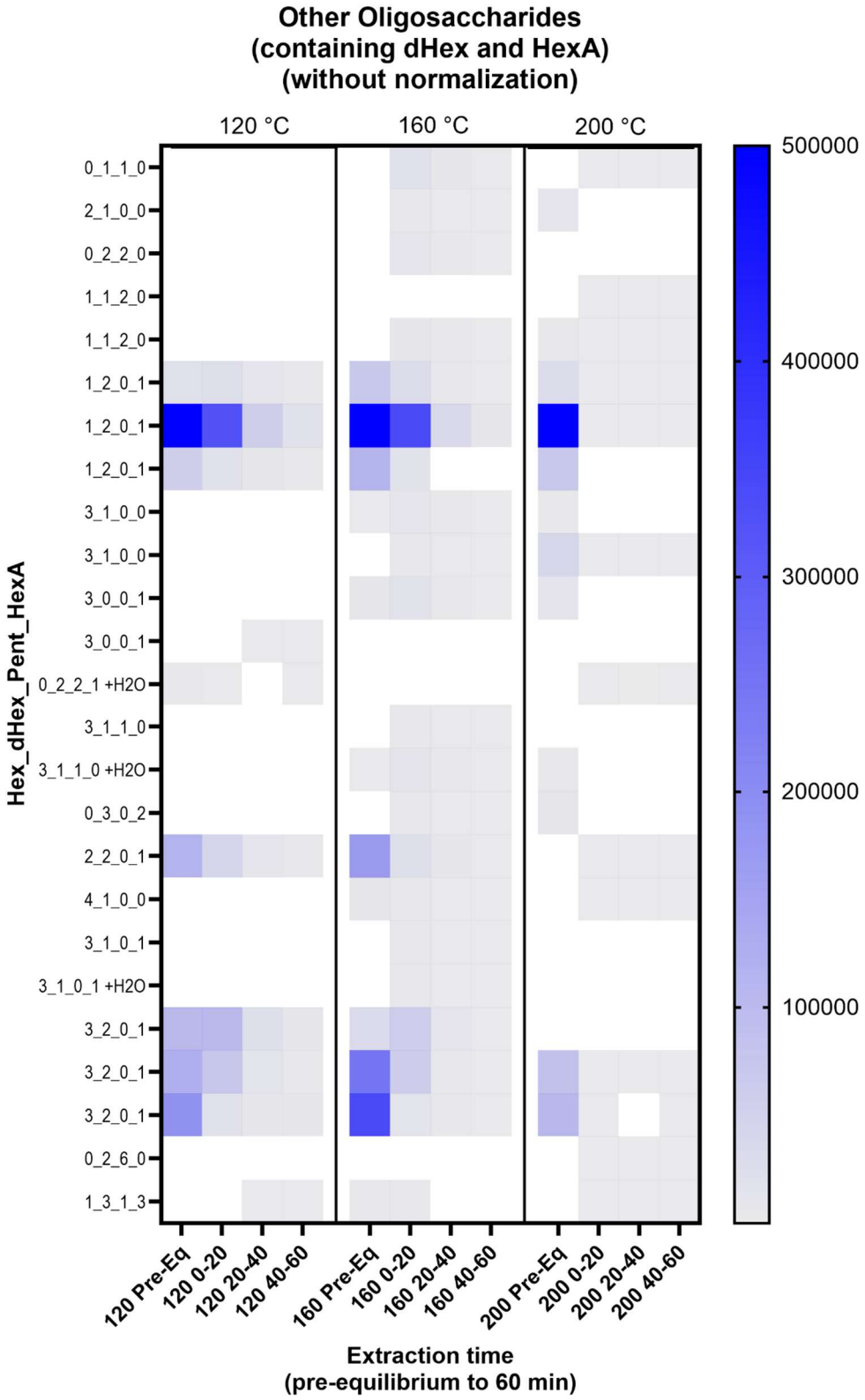
B)


Pentose-only Oligosaccharides
(without normalization)
120 °C
160 °C
200 °C
Hex_dHex_Pent_HexA
0_0_3_0
0_0_4_0
0_0_5_0
0_0_6_0
0_0_7_0
0_0_8_0
0_0_9_0
0_0_12_0
0_0_15_0
500000
400000
300000
200000
100000
120 Pre-Eq
120 0-20
120 20-40
120 40-60
160 Pre-Eq
160 0-20
160 20-40
160 40-60
200 Pre-Eq
200 0-20
200 20-40
200 40-60
Extraction time
(pre-equilibrium to 60 min)

C)


Hexose-pentose Oligosaccharides
(without normalization)
120 °C
160 °C
200 °C
Hex_dHex_Pent_HexA
2_0_1_0
2_0_1_0
2_0_1_0
2_0_1_0
2_0_1_0
1_0_3_0
1_0_3_0
2_0_2_0
2_0_2_0
2_0_2_0
2_0_2_0
2_0_2_0
2_0_2_0
2_0_2_0
3_0_1_0
3_0_1_0
3_0_1_0
3_0_1_0
3_0_1_0
3_0_1_0
3_0_1_0
3_0_1_0
3_0_1_0 +H2O
1_0_4_0
1_0_4_0
2_0_3_0
2_0_3_0
3_0_2_0
3_0_2_0
3_0_2_0
3_0_2_0
4_0_1_0
4_0_1_0
4_0_1_0
4_0_1_0
2_0_4_0
2_0_4_0
3_0_3_0
3_0_3_0
3_0_3_0
3_0_3_0
4_0_2_0
5_0_1_0
5_0_1_0
5_0_1_0
5_0_1_0
3_0_4_0
4_0_3_0
4_0_3_0
4_0_3_0
4_0_3_0
4_0_3_0 +H2O
50000
40000
30000
20000
10000
120 Pre-Eq
120 0-20
120 20-40
120 40-60
160 Pre-Eq
160 0-20
160 20-40
160 40-60
200 Pre-Eq
200 0-20
200 20-40
200 40-60
Extraction time
(pre-equilibrium to 60 min)

D)


Other Oligosaccharides
(containing dHex and HexA)
(without normalization)
120 °C
160 °C
200 °C
Hex_dHex_Pent_HexA
0_1_1_0
2_1_0_0
0_2_2_0
1_1_2_0
1_1_2_0
1_2_0_1
1_2_0_1
1_2_0_1
3_1_0_0
3_1_0_0
3_0_0_1
3_0_0_1
0_2_2_1 +H2O
3_1_1_0
3_1_1_0 +H2O
0_3_0_2
2_2_0_1
4_1_0_0
3_1_0_1
3_1_0_1 +H2O
3_2_0_1
3_2_0_1
3_2_0_1
0_2_6_0
1_3_1_3
120 Pre-Eq
120 0-20
120 20-40
120 40-60
160 Pre-Eq
160 0-20
160 20-40
160 40-60
200 Pre-Eq
200 0-20
200 20-40
200 40-60
Extraction time
(pre-equilibrium to 60 min)
500000
400000
300000
200000
100000

Supplementary Figures S2.1 A-D. Glycoprofile heatmaps of subcritical water extractions, without soluble carbohydrate content normalization. Oligosaccharides are categorized as containing hexose-only (A), pentose-only (B), hexose-pentose (C), and all others containing dHex and/or HexA (D). Individual oligosaccharides are reported on each row of the heatmaps, and the four-digit codes along the left y-axis indicate the number of monosaccharides making up the oligosaccharide (Hex_dHex_Pent_HexA). The right color legend indicates peak areas. Oligosaccharides for each subcritical temperature (120 °C, 160 °C, 200 °C) and extraction time (pre-equilibrium, 0 - 20 min, 20 - 40 min, and 40 - 60 min) are reported.

References

Almeida, J. C. de, Perassolo, M. S., Camargo, J. L., Bragagnolo, N., and Gross, J. L. (2006). Fatty acid composition and cholesterol content of beef and chicken meat in Southern Brazil. *Revista Brasileira de Ciências Farmacêuticas*, *42*(1). https://doi.org/10.1590/S1516-93322006000100012

Anderson, J. W., Gustafson, N. J., Spencer, D. B., Tietyen, J., and Bryant, C. A. (1990). Serum lipid response of hypercholesterolemic men to single and divided doses of canned beans. *The American Journal of Clinical Nutrition*, *51*(6), 1013–1019. https://doi.org/10.1093/ajcn/51.6.1013

Berg, T., Singh, J., Hardacre, A., and Boland, M. J. (2012). The role of cotyledon cell structure during in vitro digestion of starch in navy beans. *Carbohydrate Polymers*, *87*(2), 1678–1688. https://doi.org/10.1016/j.carbpol.2011.09.075

Berrios, J. D. J., Swanson, B. G., and Adeline Cheong, W. (1999). Physico-chemical characterization of stored black beans (Phaseolus vulgaris L.). *Food Research International*, *32*(10), 669–676. https://doi.org/10.1016/S0963-9969(99)00144-1

Boye, J., Zare, F., and Pletch, A. (2010). Pulse proteins: Processing, characterization, functional properties and applications in food and feed. *Food Research International*, *43*(2), 414–431. https://doi.org/10.1016/j.foodres.2009.09.003

Câmara, C., Urrea, C., and Schlegel, V. (2013). Pinto Beans (Phaseolus vulgaris L.) as a Functional Food: Implications on Human Health. *Agriculture*, *3*(1), 90–111. https://doi.org/10.3390/agriculture3010090

Campbell, K. A., and Glatz, C. E. (2009). Mechanisms of aqueous extraction of soybean oil. *Journal of Agricultural and Food Chemistry*, *57*(22), 10904–10912. https://doi.org/10.1021/jf902298a

Chang, Y.-W., Alli, I., Konishi, Y., and Ziomek, E. (2011). Characterization of protein fractions from chickpea (Cicer arietinum L.) and oat (Avena sativa L.) seeds using proteomic techniques. *Food Research International*. https://doi.org/10.1016/j.foodres.2011.08.001

Chiou, T.-Y., Neoh, T. L., Kobayashi, T., and Adachi, S. (2011). Antioxidative ability of defatted rice bran extract obtained by subcritical water extraction in bulk oil and aqueous dispersion systems. *Japan Journal of Food Engineering*, *12*(4), 147–154. https://doi.org/10.11301/jsfe.12.147

De La Fuente, M., Borrajo, A., Bermúdez, J., Lores, M., Alonso, J., López, M., … Alvarez, G. (2011). 2-DE-based proteomic analysis of common bean (Phaseolus vulgaris L.) seeds. *Journal of Proteomics*, *74*(2), 262–267. https://doi.org/10.1016/j.jprot.2010.10.004

Deb-Choudhury, S., Cooney, J., Brewster, D., Clerens, S., Knowles, S. O., Farouk, M. M., … Dyer, J. M. (2021). The effects of blanching on composition and modification of proteins in navy beans (Phaseolus vulgaris). *Food Chemistry*, *346*, 128950. https://doi.org/10.1016/j.foodchem.2020.128950

Deshpande, S. S., and Damodaran, S. (1989). Structure-Digestibility Relationship of Legume 7S

Proteins. *Journal of Food Science*, *54*(1), 108–113. https://doi.org/10.1111/j.1365-2621.1989.tb08579.x

de Almeida Costa, G. E., da Silva Queiroz-Monici, K., Pissini Machado Reis, S. M., and de Oliveira, A. C. (2006). Chemical composition, dietary fibre and resistant starch contents of raw and cooked pea, common bean, chickpea and lentil legumes. *Food Chemistry*, *94*(3), 327–330. https://doi.org/10.1016/j.foodchem.2004.11.020

de Moura, J. M. L. N., Campbell, K., Mahfuz, A., Jung, S., Glatz, C. E., and Johnson, L. (2008). Enzyme-Assisted Aqueous Extraction of Oil and Protein from Soybeans and Cream De-emulsification. *Journal of the American Oil Chemists' Society*, *85*(10), 985–995. https://doi.org/10.1007/s11746-008-1282-2

Dhingra, D., Michael, M., Rajput, H., and Patil, R. T. (2012). Dietary fibre in foods: a review. *Journal of Food Science and Technology*, *49*(3), 255–266. https://doi.org/10.1007/s13197-011-0365-5

DuBois, Michel., Gilles, K. A., Hamilton, J. K., Rebers, P. A., and Smith, Fred. (1956). Colorimetric Method for Determination of Sugars and Related Substances. *Analytical Chemistry*, *28*(3), 350–356. https://doi.org/10.1021/ac60111a017

Duc, G., Agrama, H., Bao, S., Berger, J., Bourion, V., De Ron, A. M., … Zong, X. (2015). Breeding Annual Grain Legumes for Sustainable Agriculture: New Methods to Approach Complex Traits and Target New Cultivar Ideotypes. *Critical Reviews in Plant Sciences*, *34*(1–3), 381–411. https://doi.org/10.1080/07352689.2014.898469

Elango, D., Rajendran, K., Van der Laan, L., Sebastiar, S., Raigne, J., Thaiparambil, N. A., … Singh, A. K. (2022). Raffinose family oligosaccharides: friend or foe for human and plant health? *Frontiers in Plant Science*, *13*, 829118. https://doi.org/10.3389/fpls.2022.829118

FAO. (2015, November 20). Five things we learned from the launch of the International Year of Pulses. Retrieved May 3, 2023, from 2016 International Year of Pulses website: https://www.fao.org/pulses-2016/news/news-detail/en/c/345401/

Finley, J. W., Burrell, J. B., and Reeves, P. G. (2007). Pinto bean consumption changes SCFA profiles in fecal fermentations, bacterial populations of the lower bowel, and lipid profiles in blood of humans. *The Journal of Nutrition*, *137*(11), 2391–2398. https://doi.org/10.1093/jn/137.11.2391

Geil, P. B., and Anderson, J. W. (1994). Nutrition and health implications of dry beans: a review. *Journal of the American College of Nutrition*, *13*(6), 549–558. https://doi.org/10.1080/07315724.1994.10718446

Genovese, M. I., and Lajolo, F. M. (1996). EFFECT OF BEAN (Phaseolus vulgaris) ALBUMINS ON PHASEOLIN IN VITRO DIGESTIBILITY, ROLE OF TRYPSIN INHIBITORS. *Journal of Food Biochemistry*, *20*(2), 275–294. https://doi.org/10.1111/j.1745-4514.1996.tb00556.x

Green, M. M., Blankenhorn, G., and Hart, H. (1975). Which starch fraction is water-soluble, amylose or amylopectin? *Journal of Chemical Education*, *52*(11), 729. https://doi.org/10.1021/ed052p729

Guillon, F., and Champ, M. M. J. (2002). Carbohydrate fractions of legumes: uses in human

nutrition and potential for health. *The British Journal of Nutrition*, *88 Suppl 3*, S293-306. https://doi.org/10.1079/BJN2002720

Hayat, I., Ahmad, A., Masud, T., Ahmed, A., and Bashir, S. (2014). Nutritional and health perspectives of beans (Phaseolus vulgaris L.): an overview. *Critical Reviews in Food Science and Nutrition*, *54*(5), 580–592. https://doi.org/10.1080/10408398.2011.596639

Henningsson, A. M., Nyman, E. M., and Björck, I. M. (2001). Content of short-chain fatty acids in the hindgut of rats fed processed bean (Phaseolus vulgaris) flours varying in distribution and content of indigestible carbohydrates. *The British Journal of Nutrition*, *86*(3), 379–389. https://doi.org/10.1079/bjn2001423

Huang, Y.-P., Robinson, R. C., and Barile, D. (2022). Food glycomics: Dealing with unexpected degradation of oligosaccharides during sample preparation and analysis. *Journal of Food and Drug Analysis*, *30*(1), 62–76. https://doi.org/10.38212/2224-6614.3393

Ijaq, J., Malik, G., Kumar, A., Das, P. S., Meena, N., Bethi, N., … Suravajhala, P. (2019). A model to predict the function of hypothetical proteins through a nine-point classification scoring schema. *BMC Bioinformatics*, *20*(1), 14. https://doi.org/10.1186/s12859-018-2554-y

Kataoka, M., Wiboonsirikul, J., Kimura, Y., and Adachi, S. (2008). Properties of Extracts from Wheat Bran by Subcritical Water Treatment. *Food Science and Technology Research*, *14*(6), 553–556. https://doi.org/10.3136/fstr.14.553

Kimura, A., Fukuda, T., Zhang, M., Motoyama, S., Maruyama, N., and Utsumi, S. (2008). Comparison of physicochemical properties of 7S and 11S globulins from pea, fava bean, cowpea, and French bean with those of soybean--French bean 7S globulin exhibits excellent properties. *Journal of Agricultural and Food Chemistry*, *56*(21), 10273–10279. https://doi.org/10.1021/jf801721b

Kutoš, T., Golob, T., Kač, M., and Plestenjak, A. (2003). Dietary fibre content of dry and processed beans. *Food Chemistry*, *80*(2), 231–235. https://doi.org/10.1016/S0308-8146(02)00258-3

Lubec, G., Afjehi-Sadat, L., Yang, J.-W., and John, J. P. P. (2005). Searching for hypothetical proteins: theory and practice based upon original data and literature. *Progress in Neurobiology*, *77*(1–2), 90–127. https://doi.org/10.1016/j.pneurobio.2005.10.001

Luna-Vital, D. A., Mojica, L., González de Mejía, E., Mendoza, S., and Loarca-Piña, G. (2015). Biological potential of protein hydrolysates and peptides from common bean (Phaseolus vulgaris L.): A review. *Food Research International*, *76*, 39–50. https://doi.org/10.1016/j.foodres.2014.11.024

Marsolais, F., Pajak, A., Yin, F., Taylor, M., Gabriel, M., Merino, D. M., … Chapman, R. (2010). Proteomic analysis of common bean seed with storage protein deficiency reveals up-regulation of sulfur-rich proteins and starch and raffinose metabolic enzymes, and down-regulation of the secretory pathway. *Journal of Proteomics*, *73*(8), 1587–1600. https://doi.org/10.1016/j.jprot.2010.03.013

Ma, C. Y., and Harwalkar, V. R. (1984). Chemical characterization and functionality assessment of oat protein fractions. *Journal of Agricultural and Food Chemistry*, *32*(1), 144–149.

https://doi.org/10.1021/jf00121a035

Montoya, C. A., Lallès, J.-P., Beebe, S., and Leterme, P. (2010). Phaseolin diversity as a possible strategy to improve the nutritional value of common beans (Phaseolus vulgaris). *Food Research International*, *43*(2), 443–449. https://doi.org/10.1016/j.foodres.2009.09.040

Moreno, J., Altabella, T., and Chrispeels, M. J. (1990). Characterization of alpha-Amylase-Inhibitor, a Lectin-Like Protein in the Seeds of Phaseolus vulgaris. *Plant Physiology*, *92*(3), 703–709. https://doi.org/10.1104/pp.92.3.703

Nergiz, C., and Gökgöz, E. (2007). Effects of traditional cooking methods on some antinutrients and in vitro protein digestibility of dry bean varieties (Phaseolus vulgaris L.) grown in Turkey. *International Journal of Food Science and Technology*, *42*(7), 868–873. https://doi.org/10.1111/j.1365-2621.2006.01297.x

Nielsen, S. S., Deshpande, S. S., Hermodson, M. A., and Scott, M. P. (1988). Comparative digestibility of legume storage proteins. *Journal of Agricultural and Food Chemistry*, *36*(5), 896–902. https://doi.org/10.1021/jf00083a004

Park, O. J., Ekang, N., Chang, M. J., and Kim, W. K. (2004). Resistant starch supplementation influences blood lipid concentrations and glucose control in overweight subjects. *Journal of Nutritional Science and Vitaminology*, *50*(2), 93–99. https://doi.org/10.3177/jnsv.50.93

Ramirez, C. S. V., Temelli, F., and Saldaña, M. D. A. (2021). Production of pea hull soluble fiber-derived oligosaccharides using subcritical water with carboxylic acids. *The Journal of Supercritical Fluids*, 105349. https://doi.org/10.1016/j.supflu.2021.105349

Reddy, N. R., Pierson, M. D., Sathe, S. K., and Salunkhe, D. K. (1984). Chemical, nutritional and physiological aspects of dry bean carbohydrates—A review. *Food Chemistry*, *13*(1), 25–68. https://doi.org/10.1016/0308-8146(84)90026-8

Rui, X., Boye, J. I., Ribereau, S., Simpson, B. K., and Prasher, S. O. (2011). Comparative study of the composition and thermal properties of protein isolates prepared from nine Phaseolus vulgaris legume varieties. *Food Research International*, *44*(8), 2497–2504. https://doi.org/10.1016/j.foodres.2011.01.008

Sathe, S. K. (2002). Dry bean protein functionality. *Critical Reviews in Biotechnology*, *22*(2), 175–223. https://doi.org/10.1080/07388550290789487

Shpande, S. S., and Nielsen, S. S. (1987). In vitro enzymatic hydrolysis of phaseolin, the major storage protein of phaseolus vulgaris L. *Journal of Food Science*, *52*(5), 1326–1329. https://doi.org/10.1111/j.1365-2621.1987.tb14074.x

Siddiq, M., Ravi, R., Harte, J. B., and Dolan, K. D. (2010). Physical and functional characteristics of selected dry bean (Phaseolus vulgaris L.) flours. *LWT - Food Science and Technology*, *43*(2), 232–237. https://doi.org/10.1016/j.lwt.2009.07.009

Slavin, J. (2013). Fiber and prebiotics: mechanisms and health benefits. *Nutrients*, *5*(4), 1417–1435. https://doi.org/10.3390/nu5041417

Souza Almeida, F., Furlan Goncalves Dias, F., Kawazoe Sato, A. C., and Leite Nobrega de

Moura Bell, J. M. (2021). From solvent extraction to the concurrent extraction of lipids and proteins from green coffee: An eco-friendly approach to improve process feasibility. *Food and Bioproducts Processing*, *129*, 144–156. https://doi.org/10.1016/j.fbp.2021.08.004

Souza, T. S. P., Dias, F. F. G., Koblitz, M. G. B., and. de Moura Bell, J. M. L. N (2019). Aqueous and Enzymatic Extraction of Oil and Protein from Almond Cake: A Comparative Study. *Processes*, *7*(7), 472. https://doi.org/10.3390/pr7070472

Tang, C.-H., and Sun, X. (2011). Structure–physicochemical function relationships of 7S globulins (vicilins) from red bean (Phaseolus angularis) with different polypeptide constituents. *Food Hydrocolloids*, *25*(3), 536–544. https://doi.org/10.1016/j.foodhyd.2010.08.009

Torruco-Uco, J., Chel-Guerrero, L., Martínez-Ayala, A., Dávila-Ortíz, G., and Betancur-Ancona, D. (2009). Angiotensin-I converting enzyme inhibitory and antioxidant activities of protein hydrolysates from Phaseolus lunatus and Phaseolus vulgaris seeds. *LWT - Food Science and Technology*, *42*(10), 1597–1604. https://doi.org/10.1016/j.lwt.2009.06.006

U.S. Department of Agriculture and U.S. Department of Health and Human Services. (2020). *Dietary Guidelines for Americans, 2020-2025*. *9*.

United Nations. (2023, February 10). World Pulses Day. Retrieved May 2, 2023, from United Nations website: https://www.un.org/en/observances/world-pulses-day

United Nations General Assembly. (2014). *Resolution adopted by the General Assembly on 20 December 2013 . 68/231*. Retrieved from https://daccess-ods.un.org/access.nsf/Get?OpenAgent&DS=a/res/68/231&Lang=E

USA Pulses. (2018). What are Pulses. Retrieved July 10, 2023, from https://pulses.org/us/what-are-pulses/

Wiboonsirikul, J., Hata, S., Tsuno, T., Kimura, Y., and Adachi, S. (2007). Production of functional substances from black rice bran by its treatment in subcritical water. *LWT - Food Science and Technology*, *40*(10), 1732–1740. https://doi.org/10.1016/j.lwt.2007.01.003

Wiboonsirikul, J., Mori, M., Khuwijitjaru, P., and Adachi, S. (2013). Properties of Extract from Okara by Its Subcritical Water Treatment. *International Journal of Food Properties*, *16*(5), 974–982. https://doi.org/10.1080/10942912.2011.573119

Yang, J. S., Dias, F. F. G., Pham, T. T. K., Barile, D., and de Moura Bell, J. M. L. N. (2023). A sequential fractionation approach to understanding the physicochemical and functional properties of aqueous and enzyme-assisted aqueous extracted black bean proteins. *Food Hydrocolloids*. Submitted for publication.

Yang, Y., He, S., Zhang, Y., Li, X., Liu, H., Li, Q., … Sun, H. (2021). Comparison of crude prolamins from seven kidney beans (Phaseolus vulgaris L.) based on composition, structure and functionality. *Food Chemistry*, *357*, 129748. https://doi.org/10.1016/j.foodchem.2021.129748

Yu, P. (2005). Protein secondary structures (alpha-helix and beta-sheet) at a cellular level and protein fractions in relation to rumen degradation behaviours of protein: a new approach.

The British Journal of Nutrition, *94*(5), 655–665. https://doi.org/10.1079/bjn20051532

Zhang, Y., and Romero, H. M. (2020). Exploring the structure-function relationship of Great Northern and navy bean (Phaseolus vulgaris L.) protein hydrolysates: A study on the effect of enzymatic hydrolysis. International Journal of Biological Macromolecules, 162, 1516–1525. https://doi.org/10.1016/j.ijbiomac.2020.08.019

CHAPTER 3:

Characterization of lentil (*Lens culinaris Medik*) protein and carbohydrate and assessment of green extraction methods

A modified version of this chapter is in preparation for publication submission.[2]

Abstract

Lentils (*Lens culinaris Medik*) have garnered significant interest as a sustainable and nutrient-dense food source, yet a critical knowledge gap persists regarding the characterization of their proteins and carbohydrates and the influence of green extraction techniques on lentil protein and carbohydrate composition. This study aims to address these research gaps by employing advanced proteomic and glycomic techniques to unravel the untapped potential of lentils as high-value food ingredients. Proteomics analysis explored various protein compositions resulting from different green extraction techniques, including aqueous and enzyme-assisted aqueous extractions. The use of Osborne fractionation proved to be a valuable point of comparison in the development of scalable green extraction methods. Further, enzyme-assisted extraction significantly improved protein yields, and the use of proteomics shed light on the differences in protein composition, with the enzyme-assisted extract showing fewer identified proteins, compared to aqueous extraction alone, due to proteolysis by the enzyme. The proteomics analysis also revealed the absence of a known antinutritional factor, the Kunitz-type protein inhibitor, in the enzyme-assisted extract, highlighting the potential of enzymatic extraction for reducing such antinutritional factors in the final protein extracts. In the realm of glycomics, this study evaluated the use of subcritical water extraction for recovering oligosaccharides from the insoluble byproduct remaining after lentil protein extraction. Glycoprofiling revealed distinct glycoprofiles at different extraction temperatures, with the highest temperature of 200 °C offering a significant proportion of unique oligosaccharides not observed with other conditions. The observed diversity of oligosaccharides released during

[2] Dias, F. F. G., Yang, J. S., Pham, T. T. K., Barile, D., and de Moura Bell, J. M. L. N. (2023). Unveiling the contribution of Osborne protein classes to alkaline and enzymatically extracted green lentil proteins. Manuscript in preparation.

subcritical water extraction presents optimization opportunities based on desired oligosaccharide composition. Furthermore, the glycomics techniques used in this work demonstrated their value in guiding the development of green extraction methods, providing valuable insights into the effects of time and temperature during subcritical water extraction. This integrative research study addresses key knowledge gaps surrounding lentil proteins and carbohydrates, offering detailed insights into their extraction and breakdown processes with a focus on green extraction methods. The findings contribute significantly to the development of innovative and sustainable food products, aligning with the growing demand for environmentally friendly protein sources. Furthermore, the exploration of subcritical water extraction for generating prebiotic oligosaccharides from insoluble byproducts highlights a promising avenue to enhance the value chain of lentils and minimize processing waste. The integration of proteomic and glycomic approaches provides a comprehensive perspective on the potential of lentils as high-value food ingredients, paving the way for the creation of more nutritious and environmentally sustainable food systems.

Introduction

Lentils (*Lens culinaris Medik*) are an ancient pulse crop originating from southwestern Asia. In 2021, global lentil production reached 5.8 million tons, with Canada, India, and Australia being the leading producers (FAO, 2023). Lentils are recognized for their sustainability as they are easy to cultivate, require relatively less water, and grow in a variety of soils (Asakereh et al., 2010). Additionally, lentils, like other pulses, possess nitrogen fixing capabilities that enhance soil fertility and reduce the need for synthetic fertilizers, ultimately boosting overall crop yields (Matny, 2015; Reddy, 2004). Due to these benefits, lentil production was introduced in North America in the early 1980s as a complementary pulse to existing crop rotations (Siva et al., 2017). More recently, the demand for environmentally sustainable plant-based proteins has

propelled lentils to become a major food crop in Canada and the United States (Dhul et al., 2023).

Lentils are nutrient dense, containing high protein content and complex carbohydrates such as slow digesting starch and dietary fiber (Dhull et al., 2023; Joshi et al., 2017). With a protein content of 26-28% (dry basis), lentils are not only cost-effective but also serve as an important protein source worldwide (Samaranayaka, 2017). Compared to other pulses, lentils exhibit a favorable protein and carbohydrate balance, offering higher quantities of protein, carbohydrates, and dietary fiber than most pulses (U.S. Department of Agriculture, Agricultural Research Service, 2019).

In recent years, the growing demand for sustainable protein sources and the diversification of plant protein products in the market have spurred the increased production of lentil flour and lentil protein-based food products (Romano et al., 2021). Lentil proteins have been successfully incorporated in a wide range of foods including salad dressings (Ma et al., 2016), mayonnaise (Armaforte et al., 2021), bakery products (Eckert et al., 2018; Jarpa-Parra et al., 2017), and non-dairy yogurt (Boeck et al., 2021) for both functional and nutritional purposes. While the utilization of lentils in bakery and extruded products has gained popularity, long-term success in these applications necessitates an evaluation and understanding of the effects of processing on lentils flours, the resulting functional properties, and their impact on the nutritional and sensory characteristics of the final products (Sidhu et al., 2022).

Lentil composition

Lentils are primarily made up of carbohydrates which make up about 63% of the dry matter, and protein which make up about 25% of the dry matter (U.S. Department of Agriculture, Agricultural Research Service, 2019). The proximate composition of the lentil flour utilized in this work was

24.1 ± 0.1% protein, 1.85 ± 0.14% oil, 2.34 ± 0.03% ash, 8.34 ± 0.96% moisture, and 63.4 ± 0.9% carbohydrate (by difference) (Dias et al., 2023).

Lentil proteins

The high protein content of lentils makes it a valuable protein source globally. With approximately 25% protein content, lentils are comparable to meat in terms of protein levels (Almeida et al., 2006; U.S. Department of Agriculture, Agricultural Research Service, 2019). While oilseed proteins, particularly soybeans, have been extensively studied and utilized in food products, the characterization of lentil proteins has trailed behind (Joshi et al., 2017).

Lentil proteins primarily consist of storage proteins, with smaller amounts of structural and metabolic proteins, and protective proteins (Jarpa-Parra, 2018; Luna-Vital et al., 2015). Based on their solubility, lentil proteins can be categorized into two main types: albumins and globulins. Additionally, glutelins and prolamins are present in smaller quantities (Hall et al., 2017). Globulins constitute more than 50% of total lentil protein abundance while albumins make up about 17% , glutelins account for 11%, and prolamins make up 3% (Cai et al., 2002; Hall et al., 2017).

Globulins, which are salt-soluble storage proteins, are the most abundant class of proteins in lentils (Boye et al., 2010). They can be further classified as legumin (11S) and vicilin (7S) based on the number of subunits making up the protein (Kimura et al., 2008). Albumins, on the other hand, are water-soluble and encompass various enzymatic proteins, protease inhibitors, amylase inhibitors, and lectins (Boye et al., 2010). Glutelins, soluble in dilute alkaline solutions, contain higher concentrations of methionine and cystine compared to globulins (Boye et al., 2010; Osborne, 1924). These proteins are of particular nutritional interest due to the relatively low content of essential sulfur-containing amino acids in legumes, often requiring

complementary grain proteins to create a complete protein profile (Sathe, 2002). Prolamins, which are soluble in alcohol, have a high proportion of proline and glutamine (Boye et al., 2010). This study focuses on the composition of these individual protein classes, collectively referred to as Osborne fractions, in lentils to gain insights into the effects of extraction media on the composition of extracted proteins (Osborne, 1924).

Processing lentils is necessary to enhance their nutritional quality and digestibility. However, certain processing methods such as boiling, autoclaving, or microwaving have been observed to cause decreases in protein content (Hall et al., 2017). Conversely, enzyme-assisted extraction processes (EAEP) can improve protein yields by enhancing protein solubility, therefore increasing their extractability into the aqueous media (Dias et al., 2023). Protease-assisted extraction has been studied for a variety of crops including sunflower, rice, chickpea, and soy (Campbell et al., 2016; de Moura et al., 2008; Hanmoungjai et al., 2002; Machida et al., 2022), but its application to lentils, with a specific focus on protein extractability and composition, remains largely unexplored. Previous research has characterized lentil protein albumins, globulins, and glutelins (Chang et al., 2023; Ghumman et al., 2016; Osemwota et al., 2022), but the fractions were not directly compared to protein concentrates generated using extraction methods of commercial interest.

Lentil carbohydrates

The carbohydrates present in lentils, including starch, soluble sugars, non-starch polysaccharides (dietary fiber), resistant starch (RS), and non-digestible oligosaccharides, contribute to the nutritional and health benefits associated with lentil consumption.

Starch, soluble sugars, and non-starch polysaccharides make up the majority of lentil carbohydrates (Berrios et al., 2010; Bravo et al., 1998). Various health benefits are attributed to

these components. Pulse starch is characterized by its slow glucose release and low glycemic index, and plays a role in regulating blood sugar levels (Rizkalla et al., 2002). Dietary fiber, which includes non-starch polysaccharides and oligosaccharides, is crucial for gastrointestinal health and offers various health benefits such as improved bowel health and cholesterol management (Marlett et al., 2002).

Soluble sugars, although present in relatively small amounts in lentils, also play an important nutritional role. This fraction includes mono-, di-, and small oligosaccharides, accounting for approximately 6% by weight in lentils (Hall et al., 2017). Monosaccharides, such as ribose, fructose, glucose, and galactose, make up less than 1% of seed weight (Chilomer et al., 2010; Tahir et al., 2011). Disaccharides present in lentils include sucrose and maltose (Berrios et al., 2010). Lentils are also rich in indigestible oligosaccharides known as raffinose family oligosaccharides which are discussed in further detail.

Lentils are recognized as important sources of dietary fiber. Unlike other macromolecules found in food (protein, carbohydrate, or fat) which are broken down and absorbed by the body, fiber is indigestible, leading to a number of health benefits. Consumption of dietary fiber has been linked to improved bowel health, cholesterol management, and blood glucose regulation (Berrios et al., 2010).

Lentils contain a total dietary fiber content ranging from 7% to 23%, depending on the cultivar and growth environment (Hall et al., 2017). Dietary fiber consists of various compounds such as non-starch polysaccharides and oligosaccharides (Marlett et al., 2002), and can be categorized as soluble and insoluble fiber. Insoluble fiber, including lignin, cellulose, and hemicellulose, constitutes the majority (11-19%) of total dietary fiber, while soluble fiber, primarily pectin, makes up a smaller portion (1-7%) (de Almeida Costa et al., 2006; Dhingra et al., 2012; Karaca,

et al., 2011; Silva-Cristobal et al., 2010; N. Wang and Daun, 2006). The abundance of polysaccharides making up dietary fiber can be approximated by quantifying constituent monosaccharides released through acid hydrolysis. Galacturonic acid content of lentils range between 15.6%-18.4% (Brummer et al., 2015), confirming the significant contribution of pectic polysaccharides to soluble fiber. Similarly to all pulses, glucose is the most abundant monosaccharide from insoluble fiber, indicating that cellulose is the predominant polysaccharide component (Brummer et al., 2015).

Resistant starch is another type of low-digestible carbohydrate found in lentils (Siva et al., 2018). Lentils are considered a rich source of resistant starch, with reported values of 11.4-14.9%, depending on the cultivar (Perera et al., 2010). Unlike conventional starches, resistant starch resists digestion in the small intestine and undergoes fermentation by microorganisms in the large intestine, resulting in bacterial production of short-chain fatty acids (Finley et al., 2007). It has been reported that regardless of the processing method used, the resistant starch content in processed lentils remains at relatively high levels, unlike other foods such as cereal and potato (Tovar and Melito, 1996).

The other major portion of carbohydrates are non-digestible oligosaccharides (Siva et al., 2018). These oligosaccharides include α-galactosides, such as the raffinose family oligosaccharides. Raffinose family oligosaccharides are sucrose-based carbohydrates consist of linear chains of galactose linked to the glucose residue of sucrose with α-1-6 glycosidic linkages (Avigad and Dey, 1997; Berrios et al., 2010). Raffinose family oligosaccharides serve as an energy source for the plant during seed germination, and are associated with plant stress responses and cold acclimation (Gilmour et al., 2000; Obendorf, 1997). Raffinose family oligosaccharides are well studied in pulses, and they include raffinose, stachyose and verbascose which are present in lentils in 1.6 to 2.4%, 1.7-2.9%, and 1.2-1.9%, respectively (Tahir et al., 2011).

Similarly to resistant starch, these oligosaccharides are not digested in the human small intestine and are fermented by anaerobic bacteria in the large intestine, resulting in the production of gasses (carbon dioxide, hydrogen, and methane) and short-chain fatty acids (Finley et al., 2007; Price et al., 1988). The production of these gasses can lead to digestive discomfort, flatulence, abdominal pain, and even diarrhea (Fleming, 1981; Price et al., 1988). Therefore, reducing the concentrations of these oligosaccharides to promote lentil consumption is highly desirable. Various processing methods have been shown to significantly decrease the levels of raffinose family oligosaccharides in lentils. Many studies show a considerable decrease in these oligosaccharides after soaking (Ogun et al., 1989), cooking (Verde et al., 1992; Vidal-Valverde et al., 1993), and enzymatic treatment (Price et al., 1988). Additionally, studies demonstrate that soaking pulses in alkaline solution has shown to decrease concentrations of raffinose family oligosaccharides (Abdel-Gawad, 1993).

Protein and carbohydrate analysis in the presented work with lentils

Proteomics was utilized to investigate the differences in protein composition that stem from various extraction methods and media. Due to the strong dependence of protein extraction on the specific method employed (Agrawal et al., 2013), it is challenging to compare the composition of the Osborne fractions with lentil protein extracts reported in different studies. Therefore, our approach involved sequentially fractionating lentil proteins based on solubility into Osborne fractions and assessing protein composition in parallel with lentil extracts generated using alkaline and enzymatic extraction methods. Proteomics analysis is used to investigate the differences in protein composition that stem from these various extraction methods.

Additionally, this work uses glycomics techniques to evaluate the breakdown of an insoluble byproduct remaining after enzyme-assisted extraction of protein. The green processing method investigated in this work is subcritical water extraction. Following extraction of lentil flour using an enzyme-assisted extraction process, an insoluble fraction is obtained. After removing starch from this insoluble fraction, an uncharacterized byproduct remains. Subcritical water extraction was employed to break down this byproduct to generate potentially prebiotic oligosaccharides from what would typically be considered waste. Glycomics is used to identify the various oligosaccharides released by subcritical water extraction

Experimental procedure

Figure 3.1 provides an overview of the experimental procedures involved in the analysis of lentil products, including extraction and analysis methods. Methods relevant to the protein analysis portion of this work are in the top half of the figure, and methods relevant to the carbohydrate analysis of this work are in the bottom half of the figure.

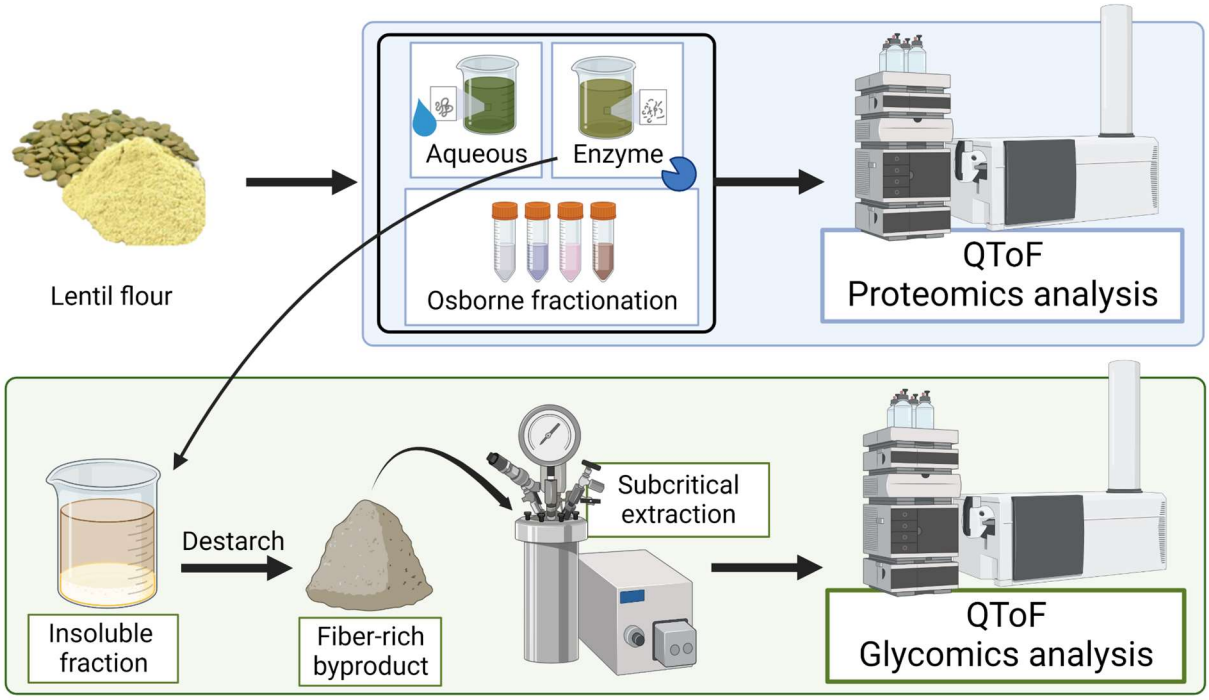


Figure 3.1. Overview of lentil protein (top) and carbohydrate (bottom) extraction and analysis. Created with BioRender.com.

Extraction methods

Prior to proteomics analysis, lentil flour was extracted using aqueous extraction (AEP), enzyme-assisted aqueous extraction (EAEP). These methods are as described by Dias et al. (2023). For the AEP, lentil flour was extracted in water at a 1:10 solids-to-liquid ratio under alkaline conditions (pH 9.0) at 50 ºC for 1 h. For the EAEP, a commercial protease (FoodPro® Alkaline Protease, Danisco, Rochester, NY, USA) was added to the extraction slurry (0.5% w/w; g enzyme/g LF). Following extraction, the slurry was centrifuged (4000 x g, 30 min, 4 °C; Allegra X-14 R, Beckman Coulter, Brea, CA, USA) to separate the protein extract (supernatant) from the fiber-rich insoluble fraction. Total protein extractability (TPE) was calculated using Equation 1 based on the amount of protein that was not extracted (present in the final insoluble fraction).

Osborne fractionation was used to separate protein fractions based on solubility in different extraction media (Osborne, 1924). Lentil flour was sequentially extracted with water, 1 M NaCl

solution, 70% ethanol, and 0.05 M NaOH at 25 °C for 1 h in each solvent (1:10 solids-to-liquid ratio) with constant stirring (120 rpm). Protein fractions at each stage in the sequence were decanted following centrifugation (4000 x g, 10 min, 4 °C), and the precipitate was resuspended in the subsequent solvent. The sequential extraction process generated the Osborne fractions: albumin-rich (ALB), globulin-rich (GLO), prolamin-rich (PRO), and glutelin-rich (GLU). The inclusion of the suffix "rich" acknowledges potential co-extraction of various protein classes in the fractions as no further purification was performed following extraction. The sequential extraction process was performed in triplicate and TPE was determined using Equation 1.

Equation 1

$$TPE\ (\%) = [1 - (\frac{Protein\ (g)\ in\ the\ Osborne\ fraction\ or\ final\ insoluble\ fraction)}{Protein\ (g)\ in\ black\ bean\ flour}) \times 100$$

SDS-PAGE

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed to visualize the molecular weight profile of the Osborne fractions and protein extracts. A pre-cast 4-20% Criterion™ TGX Precast Midi Protein Gel was used (Bio-Rad, Hercules, CA, USA). Gel images were captured using a Bio-Rad Gel Doc™ EZ Imager (Bio-Rad, Hercules, CA, USA).

Proteomics of Osborne fractions, AEP, EAEP

Protein precipitation

Extraction replicates were pooled for proteomics analysis. Aside from the prolamin fraction samples which was already extracted under organic conditions, all samples were subjected to protein precipitation as follows:

The pooled protein-extracted samples (400 µL) were mixed with ice-cold trichloroacetic acid (MilliporeSigma, St. Louis, MO)/acetone (Fisher Scientific, Waltham, MA) (10/90 v/v) (1.6 mL)

containing 20 mM dithiothreitol (Fisher Scientific, Waltham, MA) added immediately before use. Samples were incubated at -20 °C for 1.5 h, then centrifuged (4,700 rpm, 4 °C, 30 min) (Allegra X-30R Centrifuge, Beckman Coulter, Brea, CA). The supernatant was carefully discarded without disturbing the pellet. The pellet was washed with ice-cold acetone (500 µL), then centrifuged as before. Acetone washes were repeated twice more in the same manner. Following the final removal of supernatant, pellets were left to dry completely of acetone.

Reduction, alkylation, and digestion

Pelleted protein samples were dissolved in 50 mM ammonium bicarbonate (Spectrum Chemical, Gardena, CA) (250 µL). For prolamin fraction samples, ammonium bicarbonate was added to the sample for a final concentration of 50 mM. The equivalent of 50 µg protein (determined by Qubit 3.0 fluorometer assay, ThermoFisher Scientific, Waltham, MA) was transferred to fresh tubes for digestion. Concentrated samples were diluted to 1 µg/µL protein using additional 50 mM ammonium bicarbonate. Dithiothreitol was added to a final concentration of 5 mM, and samples were incubated (60 °C, 30 min). Iodoacetamide (MilliporeSigma, St. Louis, MO) was added to a final concentration of 20 mM, and samples were incubated in the dark (room temperature, 30 min). Trypsin (Promega, Madison, WI) was added (1:50 ratio), and samples were incubated (37 °C, overnight) (Eppendorf ThermoMixer C, Enfield, CT). The following day, trypsin was inactivated by reducing pH to 2-3 using 1% TFA. All samples were centrifuged to pellet insoluble material (14,000 xg, 4 °C, 15 min) (Eppendorf Centrifuge 5424, Enfield, CT).

C18 Cleanup

Digested samples were further purified by microplate C18 SPE (Glygen, Columbia, MD). Acetonitrile (LC-MS grade, Fisher Scientific, Waltham, MA) and trifluoroacetic acid (TFA) (MilliporeSigma, St. Louis, MO) were used for C18 SPE. The microplate wells were activated

with 99.9% acetonitrile/0.1% TFA (v/v) and equilibrated with 1% acetonitrile /0.1% TFA in water (v/v/v). Samples were loaded, and wells were washed with 1% acetonitrile /0.1% TFA in water (v/v/v) (1.2 mL). Peptides were eluted with 80% acetonitrile /0.1% TFA in water (v/v/v) (600 µL). The eluent was dried (Eppendorf Vacufuge plus, Enfield, CT) and redissolved in 3% acetonitrile /0.1% FA in water (v/v/v) for LC-MS/MS analysis.

LC-MS/MS analysis

LC-MS/MS analysis was performed on an Agilent 6520 Accurate-Mass Q-TOF LC-MS with a Chip Cube interface (Agilent Technologies, Santa Clara, CA). Mobile phases contained MilliQ water, acetonitrile (LC-MS grade, Fisher Scientific, Waltham, MA) and formic acid (LC-MS grade, Fisher Scientific, Waltham, MA). Samples were injected onto a Zorbax 300SB-C18, 5 µm, 150 mm x 75 µm Chip with 0.1% FA in water (v/v) at a flow rate of 4 µL min−1. Chromatographic separation of peptides was performed with a gradient consisting of 0.1% FA in water (v/v; A) and 90% acetonitrile, 0.1% FA in water (v/v/v, B) at a flow rate of 0.3 µL min−1. The 80-min gradient was ramped from 0-30% B, 5-60 min; 30-100% B, 60-65 min; 100% B, 65-70 min; 100-0% B, 70 min; 0% B, 70-80 min. The capillary voltage was set to 1950 V. The drying gas was set to 325 °C with a flow rate of 5 L min−1. Mass-to-charge ratio (m/z) was scanned at a rate of 8 spectra sec^{-1} in the m/z range of 275-1700. The precursors were selected based on abundance and isolated with a width of 1.3 m/z for fragmentation. A ramped collision energy with the equation (0.03 × m/z + 2) was applied to ions of any charges. The MS/MS analysis was scanned at a rate of 0.63 spectra sec^{-1} in the m/z range of 50–1750.

Data analysis and protein identification PEAKS Studio X+ (Bioinformatics Solutions Inc., Waterloo, ON, Canada) was used for peptide and protein identification. Peptides were identified through database search using NCBI databases with species names *Lens culinaris*, *Pisum sativum*, *Cicer arietinum*, and *Vicia faba* (https://www.ncbi.nlm.nih.gov/, accessed 2/11/2022). The mass error tolerance was 20 ppm and 0.035 Da for the precursor and fragment ions,

respectively. The enzyme was set to “Trypsin” with a specific digestion mode. The number of maximum missed cleavages per peptide was set to 2. Carbamidomethylation was set as a fixed PTM. A maximum of 5 variable modifications, including oxidation, phosphorylation, and deamidation were allowed. The results were filtered with a false discovery rate of 1.0%. Only proteins with at least 1 unique peptide were retained. Database matches were manually inspected to select the correct protein match.

Glycoprofiling of subcritical samples

Prior to subcritical water extraction, the insoluble fraction remaining after protein extraction underwent destarching (1:10 SLR, pH 6.0, 90 °C, 2 h) using commercial amylase (1% w/w FoodPro® AHT). To assess the impact of temperature on glycoprofiles, subcritical water extraction was conducted at three temperatures (120 °C, 160 °C, and 200 °C). To assess the effects of extraction time over a 60 min extraction, fractions were collected for pre-equilibrium (the time after reaching the set temperature but before reaching the target pressure), 0-20 min, 20-40 min, and 40-60 min.

Oligosaccharides released from subcritical extraction of the insoluble fraction were isolated by removing proteins via cold ethanol precipitation. Two volumes of cold ethanol were added to reconstituted fractions and stored at -20°C for 1 h. Samples were centrifuged (4200 xg, 4 °C, 30 min) (Allegra X-30R Centrifuge, Beckman Coulter, Brea, CA). Supernatants were dried by centrifugal evaporation (Genevac miVac, Ipswich, United Kingdom) then reconstituted on water for two stages of solid-phase extraction (SPE). For carbohydrate-normalized analysis, sample volumes were normalized to total soluble carbohydrate content as determined by the phenol-sulfuric acid method (DuBois et al., 1956) to give 20 mg/mL soluble carbohydrate concentration. Samples were cleaned by C18 microplate (Glygen, Columbia, MD), followed by microplate porous graphitized carbon (PGC) SPE (Thermo Fisher Scientific, Waltham, MA). Prior to sample

loading onto SPE microplates, reconstituted samples were centrifuged at 14,000 x g, 4°C, for 30 minutes (Eppendorf Centrifuge 5424, Enfield, CT) to remove particulates.

Reconstituted samples were cleaned by SPE as follows. Between all additions, microplates were centrifuged at 1300 rpm, 20°C, for 1 min (Allegra X-30R Centrifuge, Beckman Coulter, Brea, CA). C18 microplates were conditioned with 200 µL acetonitrile (x 3), and equilibrated with 200 µL water (x 3). Samples (200 µL) were loaded atop of a fresh collection plate. The flowthrough and subsequent washes (3 x of 200 µL water) were collected for PGC SPE.

PGC microplates were conditioned with 200 µL 80/20 acetonitrile/water containing 0.1% TFA (x 5). Wells were equilibrated with 200 µL water (x 4). The collected flowthrough and washes from C18 SPE were loaded in 200 µL aliquots, centrifuging between additions. Wells were washed with 200 µL water (x 6). A fresh collection plate was used to collect the eluent (3 x 200 µL 40/60 acetonitrile/water containing 0.1% TFA). Eluents were transferred to Eppendorf tubes and dried by centrifugal evaporation (Genevac miVac, Ipswich, United Kingdom). Samples were reconstituted in MilliQ water for LC-MS analysis.

Purified oligosaccharides were analyzed by LC-MS/MS using an Agilent 6520 Accurate-Mass Q-TOF LC-MS with a Chip Cube interface coupled to an Agilent 1200 Series high performance liquid chromatography system (Agilent Technologies, Santa Clara, CA, USA). The mobile phases were composed of 0.1 % formic acid (FA), 3% acetonitrile in water (mobile phase A) and 0.1% FA, 90% acetonitrile in water (mobile phase B). Samples containing oligosaccharides were injected onto an Agilent PGC-Chip II (G4240-64010) at a flow rate of 4 µL/min. Oligosaccharides were chromatographically separated over a 60 min gradient at a flow rate of 0.3 µL/min, beginning with 100% A for 2.5 min, ramping from 0% to 16% B in 20 min, increasing from 16% to 44% B in 10 min, then 44% to 100% B in 5 min, and held at 100% B for 10 min.

The system was re-equilibrated for 15 min at 100% A prior to the next injection. The drying gas was set at 350 °C with a flow rate of 5 L/min. The electrospray ion source was in positive ion mode with a capillary voltage of 1850 V. The ions were scanned within the range of m/z 150–2500 at a rate of 1 spectrum/sec. The four most abundant ions in each MS analysis cycle were isolated for tandem MS analysis with ramped collision energy (CE; CE = 0.02 × m/z – 3.5). Reference ions m/z 922.009798 and m/z 1221.990637 were used for continual mass calibration throughout the analysis.

Fragmentation data was annotated by Glyconote (https://github.com/MingqiLiu/GlycoNote) and manually inspected using Agilent Masshunter Qualitative Analysis (B.07.00, Agilent Technologies) to identify oligosaccharides structures. Verified oligosaccharides were compiled into a novel library. Peaks were manually integrated by Agilent Masshunter Profinder (B.08.00, Agilent Technologies) using targeted feature extraction which included the monoisotopic masses (with a mass error within 20 ppm) and retention times for all identified oligosaccharides. In-source fragment ions and dimer and trimer aggregates were manually searched and summed to approximate their actual abundance (Huang et al., 2022). GraphPad Prism (ver. 9.4.0) was used to generate heatmaps from peak area data.

Results

Total protein extracted

Protein extraction yields for the alkaline extraction (AEP), enzyme-assisted extraction (EAEP), and Osborne fractionation are presented in Figure 3.2 A, B. Osborne fractionation achieved 97% total protein extractability (i.e., 3% of the total protein remained in the unextracted insoluble fraction). Of the Osborne fractions, a majority of the extracted proteins were in the albumin-rich (43%) and globulin-rich (37%) fractions, followed by the glutelin-rich (14%) and prolamin-rich (3%) fractions.

Figure 3.2A demonstrates that AEP and EAEP were highly effective in extracting proteins from lentil flour. A significant increase in the total protein extractability was observed when Alkaline Protease was added to the extraction slurry. Specifically, protein extractability increased from 81% (AEP, no enzyme) to 87% (EAEP, 0.5% w/w enzyme). Such increases have also been demonstrated for the enzymatic extraction of other pulse proteins such as chickpea (from 63 to 84%) and common bean (from 75 to 81%) using FoodPro® Alkaline Protease (Machida et al., 2022; Yang et al., 2023).

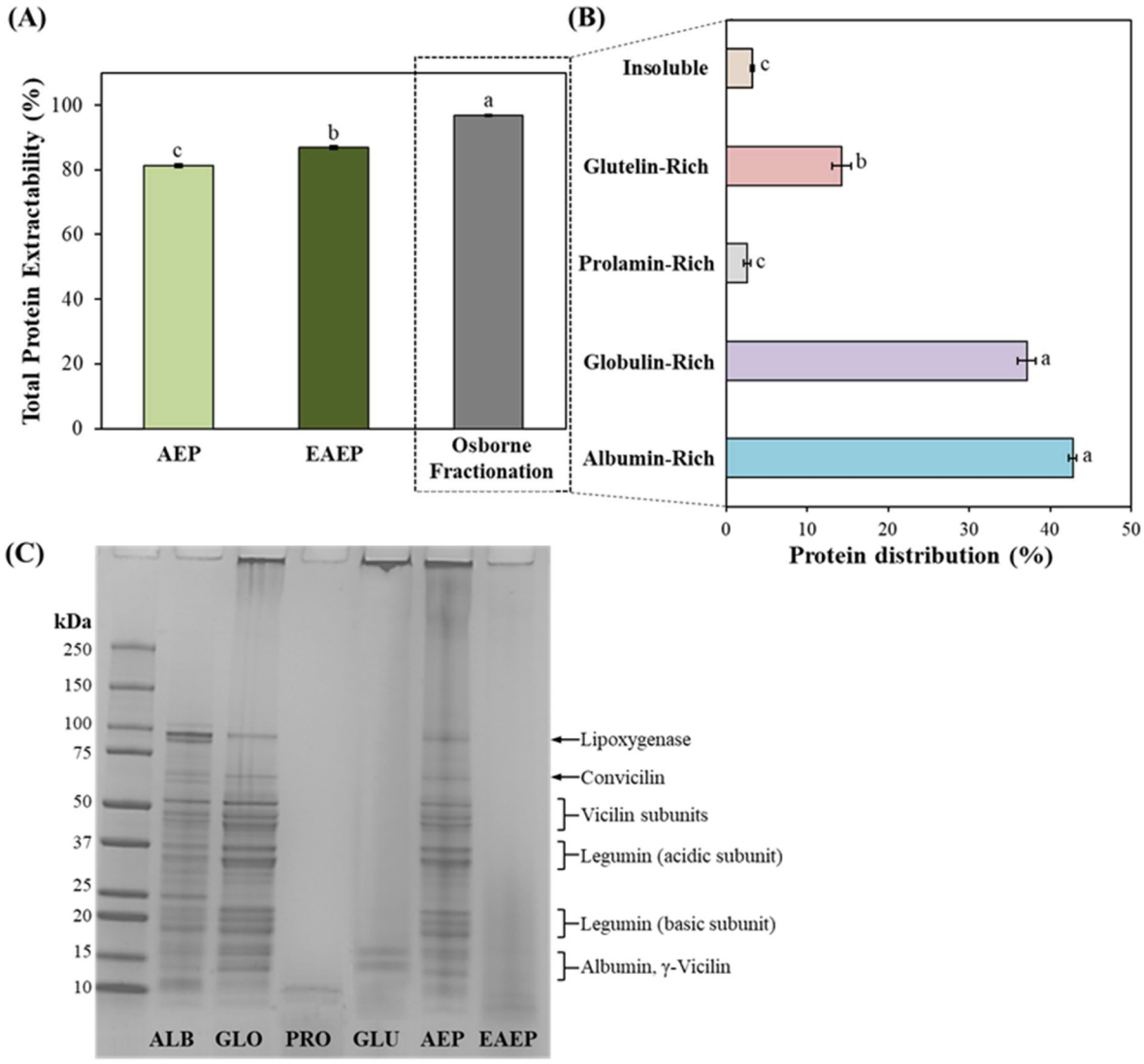


Figure 3.2. (A) Total protein extractability (%), (B) protein distribution of the Osborne fractions (%), and (C) reducing SDS-PAGE of the Osborne fractions (ALB = albumin-rich fraction, GLO = globulin-rich fraction, PRO = prolamin-rich fraction, GLU = glutelin-rich fraction), AEP (aqueous extraction process), and EAEP (enzyme-assisted aqueous extraction process). Values are reported as the mean ± SD (n = 3), and different letters indicate statistical differences by ANOVA with Tukey's test ($p < 0.05$).

SDS-PAGE

To better understand how fractionation and extraction affect the protein profile of the lentil extracts, an SDS-PAGE in the presence of a reducing agent (β-mercaptoethanol) was performed (Fig 3.2C). The protein profiles exhibited a variety of bands with molecular weights

(MWs) between 10 and 100 kDa, resembling previously reported SDS-PAGE for lentil proteins (Alonso-Miravalles et al., 2019; Shrestha et al., 2023). Overall, the major bands observed in the albumin, globulin, and AEP lanes could be attributed to the polypeptide constituents of the major storage proteins in lentils: legumin (~40 and 20 kDa subunits) and vicilin (~50 kDa). Legumin, an 11S globulin, is a hexameric protein formed by subunits with MW ~ 60 kDa, which consist of an acidic (~ 40 kDa) and a basic (~ 20 kDa) subunit linked by a disulfide bond (Shewry et al., 1995). Therefore, in reducing conditions, the bands around 37-40 and 20-25 kDa may be attributed to the acidic and basic subunits of legumin, respectively (Alonso-Miravalles et al., 2019). Vicilin is a ~150 kDa 7S trimeric globulin with subunits that are not linked by disulfide bonds (Shewry, 1995); notably, vicilins in lentils have been demonstrated to be relatively heterogeneous in MW, which may explain the many bands observed around 50 kDa (Scippa et al., 2010). The ~97 kDa and 70 kDa bands could correspond to lipoxygenase and convicilin, respectively (Chang et al., 2023; Shevkani et al., 2019).

The albumin and globulin fraction protein profiles were similar, with some changes in band intensity. These similarities indicate that globulins were co-extracted in the initial water-soluble phase. The prolamin and glutelin fractions exhibited different protein profiles with predominantly low MW peptides (< 20 kDa). The MW distribution of the AEP skim proteins was similar to the albumin and globulin fractions, which was expected given that the AEP was performed under aqueous conditions. The lack of visible distinct subunits in the EAEP lane shows that the use of enzyme during the extraction extensively hydrolyzed the major storage proteins, resulting in the generation of smaller peptides with MW < 15 kDa.

Proteomics results of AEP, EAEP, and Osborne fractions

The Osborne fractions, AEP, and EAEP extracts were analyzed using liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS) to elucidate differences in protein species

between the samples. There was a lack of genome sequencing and NCBI protein database entries specific to lentils which limited the number of spectral matches and protein identifications. To address this challenge, a cross-species search was performed to include other pulse species. As with any cross-species proteomics study, database searching of cross species introduces a higher degree of error (Wright et al., 2010). To minimize this, the identified lentil proteins with the highest sequence coverage were searched using NCBI Blastp (https://blast.ncbi.nlm.nih.gov/) to identify homologous protein sequences. The cross species presented in this work (pea, chickpea, and fava bean) showed protein homology to lentils and thus were selected for database searching. To further minimize the high degree of error associated with cross-species proteomics studies, only higher quality UniProtKB/Swiss-Prot reviewed proteins are reported.

A total of 129 proteins were identified among reviewed and unreviewed source databases (Supplementary Materials Table S3.1), of which 37 proteins were UniProtKB/Swiss-Prot reviewed, meaning they have been manually annotated and are considered higher quality identifications (summarized in Table 3.1 with full data in Supplementary Materials Table S3.1). A majority of the identified proteins were from the pea database (27), followed by fava bean (7), and lastly, lentil (3). However, as expected, the sequence coverage (%) of the protein species in Table 1 demonstrates that the proteins identified using the lentil database (i.e., lectin, Bowman-Birk protease inhibitor, lipid-transfer protein) had the highest coverage. The proteins shown in Table 1 were categorized based on functional class (e.g., storage, disease/defense, metabolism) (Ialicicco et al., 2012; Scippa et al., 2010; X. Wang et al., 2016).

Of the Osborne fractions, the albumin fraction contained the most identified proteins (32), followed by the globulin (21), glutelin (17), and prolamin (6) fractions. This suggests that many proteins (particularly those in the metabolism-related functional class) are readily extracted in

water and were therefore completely extracted in the first stage of fractionation. The relatively smaller number of identified proteins in the prolamin fraction may also reflect the conditions of proteomic sample preparation. Proteins in the prolamin fraction were extracted in high organic conditions, but proteomics sample preparation and protein digestion were performed in high aqueous conditions. As such, it is possible that key proteins in the prolamin fraction were either not captured during initial protein precipitation steps, or precipitated over the course of proteomics sample preparation and were not in solution during protein digestion. To overcome this, subsequent studies may utilize alternative proteomics sample preparation techniques such as FASP (filter aided sample preparation) to minimize differences in protein physical properties.

In comparing the proteins identified, major storage proteins such as vicilin and provicilin were detected in all fractions, and legumin and convicilin were found in all fractions for lentil proteins except prolamin. The other protein species in the extracts were disease/defense-related proteins including lectins and trypsin inhibitors (Kunitz and Bowman-Birk types). Some differences were observed between the fractions for the metabolism-related proteins, but because a majority (~70-80%) of the proteins in lentils are storage proteins (Ialicicco et al., 2012; Joshi et al., 2017), these minor differences may not be significant in the overall protein profile of the fractions.

Fewer proteins were identified in the EAEP protein extract compared to the AEP (17 for EAEP vs. 25 for AEP), which demonstrates that proteolysis degraded some of the lentil proteins prior to proteomics analysis. This is corroborated with the results observed in the SDS-PAGE showing very few intact proteins. In particular, the Kunitz-type protein inhibitor, a known antinutritional factor, was not detected in the EAEP. Antinutritional factors can reduce the metabolic utilization and/or digestion of plant foods, leading to impaired gastrointestinal functions and metabolic performance (Fekadu Gemede, 2014). These results suggest that

enzymatic extraction of lentil proteins could be an effective strategy to reduce the antinutritional factors in the resulting extracts. Interestingly, the Kunitz-type protein inhibitor was not detected in any of the Osborne fractions. This could signify that the mild heating in the AEP (50 °C) was required for the extraction of this specific protein, compared to the sequential fractionation that was performed at room temperature.

Table 3.1. UniProtKB/Swiss-Prot reviewed protein identifications in the Osborne fractions, aqueous extraction (AEP), and enzyme–assisted extraction (EAEP) lentil protein extracts by LC-MS/MS. Checkmarks (✓) denote the presence of the protein in the sample.

Description	Accession	Database	Avg. Mass (Da)	-10lgP	Coverage (%)	ALB	GLO	PRO	GLU	AEP	EAEP
Storage-Related Proteins											
Albumin-2	sp|P08688.1|ALB2_PEA	pea	26238	107.67	7	✓	✓			✓	✓
Convicilin	sp|P13915.1|CVCA_PEA	pea	66990	218.78	14	✓	✓		✓	✓	✓
Ferritin-1 (chloroplastic)	sp|P19975.2|FRI1_PEA	pea	28619	93.78	14	✓	✓			✓	✓
Legumin A	sp|P02857.1|LEGA_PEA	pea	58805	183.61	13	✓	✓		✓	✓	✓
Legumin A2	sp|P15838.1|LEGA2_PEA	pea	59270	206.81	14	✓	✓		✓	✓	✓
Legumin J	sp|P05692.1|LEGJ_PEA	pea	56895	235.73	26	✓	✓		✓	✓	✓
Provicilin	sp|P02855.1|VCLA_PEA	pea	31540	200.88	21	✓	✓	✓	✓	✓	✓
Vicilin	sp|P13918.2|VCLC_PEA	pea	52231	312.25	39	✓	✓	✓	✓	✓	✓
Disease/Defense-Related Proteins											
22.7 kDa class IV heat shock protein	sp|P19244.1|HSP41_PEA	pea	22734	80.71	11	✓			✓		✓
Bowman-Birk type proteinase inhibitor	sp|Q8W4Y8.2|IBB_LENCU	lentil	12266	221.51	35	✓	✓	✓		✓	✓
Heat shock 70 kDa protein (mitochondrial)	sp|P37900.1|HSP7M_PEA	pea	72301	71.72	3	✓					
Kunitz-type trypsin inhibitor-like 1 protein	sp|Q41015.2|PIP21_PEA	pea	23792	74.57	17					✓	
Lectin	sp|P02870.2|LEC_LENCU	lentil	30352	273.47	29	✓	✓		✓	✓	✓
Seed biotin-containing protein SBP65	sp|Q41060.1|SBP65_PEA	pea	59554	170.65	11	✓	✓	✓	✓	✓	
Superoxide dismutase [Cu-Zn]	sp|Q02610.2|SODC_PEA	pea	15323	90.11	15	✓	✓			✓	

Superoxide dismutase [Cu-Zn] (chloroplastic)	sp|P11964.1|SODCP_PEA	pea	20626	73.48	12	✓					
Metabolism-Related Proteins											
Alcohol dehydrogenase 1	sp|P12886.1|ADH1_PEA	pea	41155	150.37	14	✓	✓		✓	✓	✓
Alpha-1 4 glucan phosphorylase L isozyme (chloroplastic/amyloplastic)	sp|P53536.2|PHSL_VICFA	fava bean	113580	320.65	18	✓	✓		✓	✓	
Alpha-glucan phosphorylase (H isozyme)	sp|P53537.1|PHSH_VICFA	fava bean	95924	90.37	3	✓	✓				
Fructose-bisphosphate aldolase (cytoplasmic)	sp|P46257.1|ALF2_PEA	pea	38491	128.89	12	✓	✓			✓	
Fructose-bisphosphate aldolase 1 (chloroplastic)	sp|Q01516.1|ALFC1_PEA	pea	38657	91.53	8	✓				✓	
Glucose-1-phosphate adenylyltransferase small subunit 2 (chloroplastic)	sp|P52417.1|GLGS2_VICFA	fava bean	56060	89.61	5	✓					
Glyceraldehyde-3-phosphate dehydrogenase (cytosolic)	sp|P34922.1|G3PC_PEA	pea	36609	139.47	22	✓	✓		✓	✓	
Lipoxygenase-3	sp|P09918.1|LOX3_PEA	pea	97629	257.33	28	✓	✓		✓	✓	✓
Non-specific lipid-transfer protein 5	sp|A0AT31.1|NLTP5_LENCU	lentil	11686	191.9	34			✓			
Nucleoside diphosphate kinase 1	sp|P47922.1|NDK1_PEA	pea	16463	125.35	23	✓	✓		✓	✓	✓
Phosphoglucomutase (cytoplasmic)	sp|Q9SM60.1|PGMC_PEA	pea	63325	134.78	12	✓				✓	
Probable sucrose-phosphate synthase	sp|Q43876.1|SPSA_VICFA	fava bean	118204	107.45	2	✓					
Ribulose bisphosphate carboxylase large chain	sp|P04717.3|RBL_PEA	pea	52763	107.38	6	✓					
RuBisCO large subunit-binding protein subunit alpha (chloroplastic)	sp|P08926.2|RUBA_PEA	pea	61979	147.32	12	✓				✓	
RuBisCO large subunit-binding protein subunit beta (chloroplastic)	sp|P08927.2|RUBB_PEA	pea	62984	94.89	4	✓					

Other

14-3-3-like protein A	sp|P42653.1|1433A_VICFA	fava bean	29420	96.07	10				✓		
ATP synthase subunit alpha (mitochondrial)	sp|P05493.2|ATPAM_PEA	pea	55045	78.16	4	✓	✓				
Elongation factor 1-alpha	sp|O24534.1|EF1A_VICFA	fava bean	49244	145.02	12	✓			✓		✓
GTP-binding nuclear protein Ran/TC4	sp|P38548.1|RAN_VICFA	fava bean	25290	136.98	16	✓				✓	✓
Histone H4	sp|P62788.2|H4_PEA	pea	11409	131.2	30		✓		✓	✓	✓
Polyubiquitin	sp|P69322.2|UBIQP_PEA	pea	42699	118.04	7			✓		✓	

-10lgP: Score from Peaks Xpro software

ALB: albumin-rich fraction, GLO: globulin-rich fraction, PRO: prolamin-rich fraction, GLU: glutelin-rich fraction

AEP: aqueous extraction process protein extract, EAEP: enzyme-assisted aqueous extraction process protein extract

Glycoprofiling of subcritical water extractions

The present study investigates the subcritical water extraction of insoluble byproducts remaining after enzyme-assisted extraction process (EAEP), with a specific focus on the identification and profiling of oligosaccharides. While a few studies have explored the extraction of legume byproducts using subcritical water (Ramirez et al., 2021; Wiboonsirikul et al., 2013), the utilization of subcritical water extraction to release oligosaccharides from lentil waste following aqueous and enzymatic extraction approaches is novel. Prior to subcritical water extraction, this byproduct underwent destarching using commercial amylase, resulting in an 80% reduction in starch content. To assess the impact of temperature and time on glycoprofiles, subcritical water extraction was conducted at three temperatures (120 °C, 160 °C, and 200 °C) and fractions were collected over 60 min total extraction time (pre-equilibrium, 0-20 min, 20-40 min, and 40-60 min).

For building a library, all fractions were normalized based on their respective total soluble carbohydrate content as determined by the phenol-sulfuric acid method (DuBois et al., 1956) to ensure that an approximately equal amount of carbohydrates was analyzed by LC-MS/MS across all fractions. This normalization ensured that oligosaccharides would be adequately detected among all fractions, enabling a maximum number of oligosaccharides to be included in the library. All oligosaccharides in the library consisted of Hexose (Hex), Pentose (Pent), deoxyhexose (dHex), and acidic sugars (HexA) as possible constituent monosaccharides. All features in the lentil subcritical water extraction library are reported in Supplementary Table 3.2.

Among all subcritical water extraction temperatures and fractions, a total of 99 oligosaccharides were identified and included in the library as depicted in Figure 3.3. Oligosaccharides were classified by their monosaccharide composition as containing hexoses-only, pentose-only, hexose-pentose, and remaining others containing dHex and/or HexA. The reported results

included summed in-source fragments and dimer and trimer aggregates in order to approximate the true abundance of each oligosaccharide (Huang et al., 2022). Figure 3.4 illustrates the total counts of oligosaccharides for each subcritical water extraction fraction and temperature, classified by their monosaccharide composition. This classification allows for an assessment of the differences in oligosaccharide profiles resulting from varying subcritical water extraction temperatures.

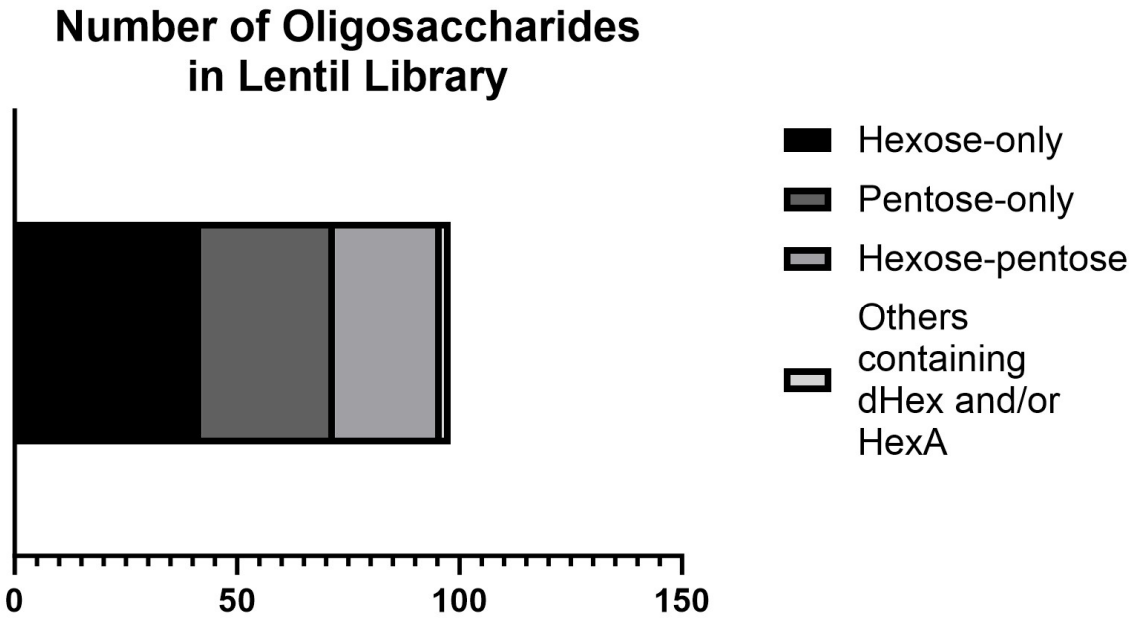


Figure 3.3. Total number of oligosaccharides in lentil library, categorized based on constituent monosaccharides. Oligosaccharides were categorized as containing only hexoses, only pentoses, hexoses and pentoses, and all others containing deoxyhexoses (dHex) and/or acidic sugars (HexA).

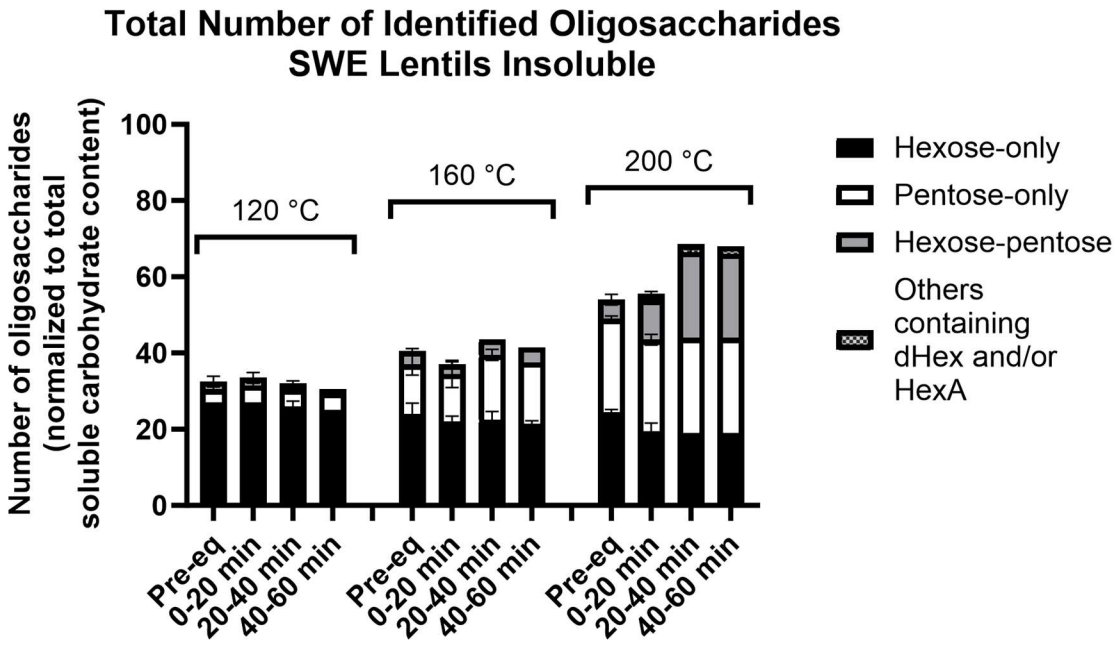


Figure 3.4. Total counts of oligosaccharides identified in subcritical water extracts. Extracts were normalized to soluble carbohydrate content for analysis. Subcritical water extraction was evaluated at three temperatures, 120 °C, 160 °C, and 200 °C. The total time of extraction was 60 min, with fractions collected during pre-equilibrium (the time for the extraction vessel to achieve set pressure), 0 - 20 min, 20 - 40 min, and 40- 60 min. At each temperature and time, oligosaccharides are categorized by their monosaccharide composition.

Distinct oligosaccharide profiles were observed for each subcritical water extraction (subcritical water extraction) temperature condition, as depicted in Figure 3.4. The highest subcritical extraction temperature of 200 °C yielded the highest number of identified oligosaccharides, particularly in the later fractions (20-40 min and 40-60 min into extraction). The intermediate extraction temperature of 160 °C exhibited an intermediate number of identified oligosaccharides. Further, the lowest subcritical extraction temperature of 120 °C resulted in the fewest identified oligosaccharides, with slightly higher counts in the early fractions (pre-equilibrium and 0-20 min). These findings are consistent with studies investigating subcritical extraction of wheat and rice bran which reported that carbohydrate content increased with increasing temperature up to 180-200 °C (Chiou et al., 2011; Kataoka et al., 2008; Wiboonsirikul et al., 2007). However, a more detailed analysis of the types of oligosaccharides extracted at

each temperature can be obtained by categorizing them based on their constituent monosaccharides.

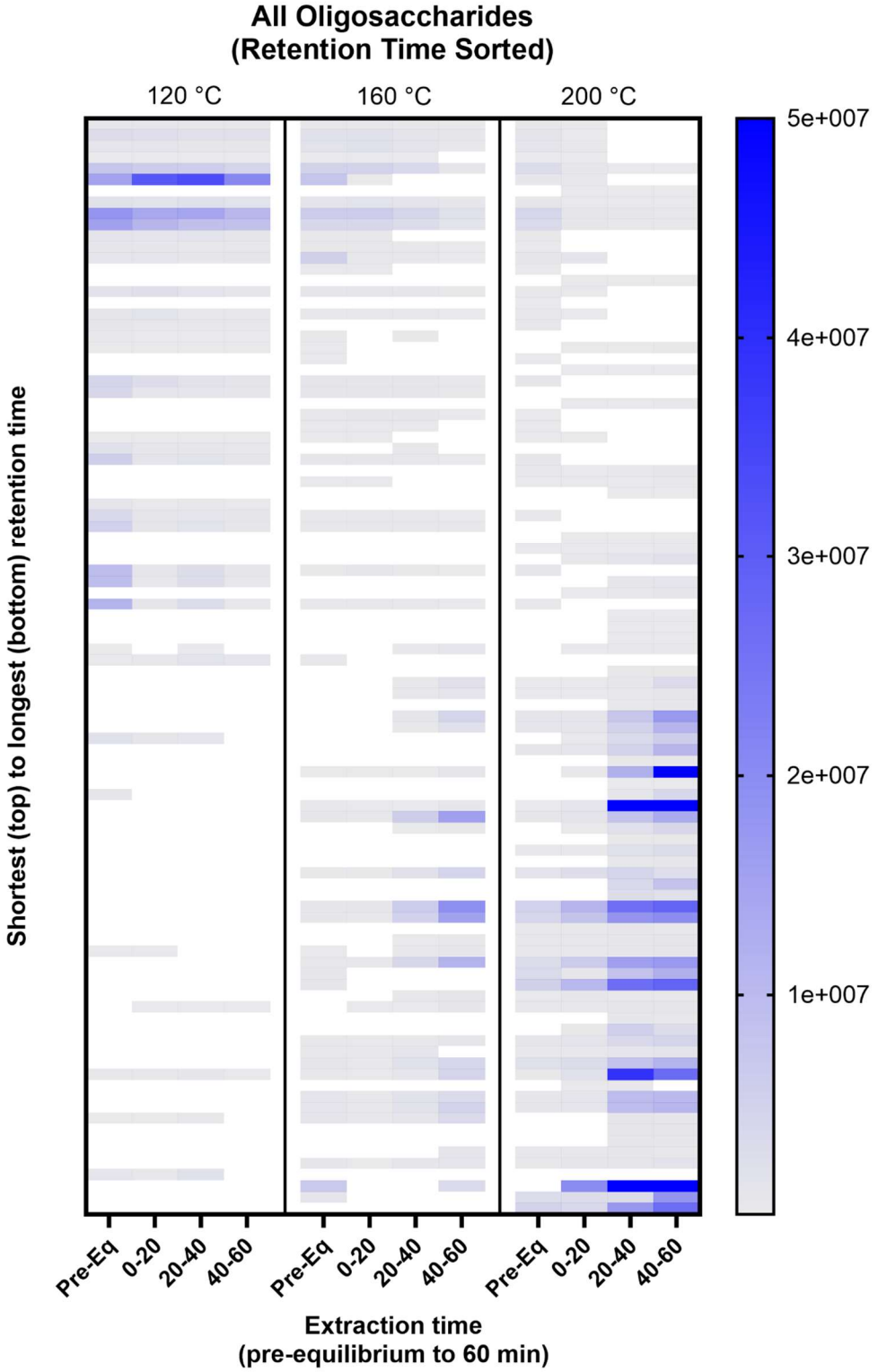
Figure 3.3 presents the total number of oligosaccharides constituting the lentil subcritical water extraction library, classified by their monosaccharide compositions. In total, 42 hexose-only oligosaccharides, 30 pentose-only oligosaccharides, 24 hexose-pentose oligosaccharides, and 2 oligosaccharides containing dHex and or HexA were identified. Figure 3.4 shows the profile of these oligosaccharides categories in relation to subcritical water extraction time and temperature. At the lowest temperature, 120 °C, the identified oligosaccharides predominantly consisted of hexoses. At the intermediate temperature, 160 °C, oligosaccharides were primarily hexose-only, with a moderate number of pentose-only oligosaccharides. At the highest temperature, 200 °C, the largest number of oligosaccharides consisted of pentose-only, with a slightly smaller count of hexose-only oligosaccharides. Most interestingly, a significant number of hexose-pentose oligosaccharides were observed in the later fractions of extraction at 200 °C. This may indicate a more complete breakdown of cell wall polysaccharides at the highest temperature, resulting in an abundance of pentose-only and hexose-pentose oligosaccharides. Specifically, in the later fractions, the counts of hexose-pentose oligosaccharides surpassed those of hexose-only oligosaccharides, which was unique to the highest extraction temperature.

To visualize the peak area data, heatmaps were generated (Figures 3.5A-D). Individual oligosaccharides are reported on each row of the heatmap, and the four-digit codes along the y-axis indicate the number of monosaccharides making up the oligosaccharide (Hex_dHex_Pent_HexA).

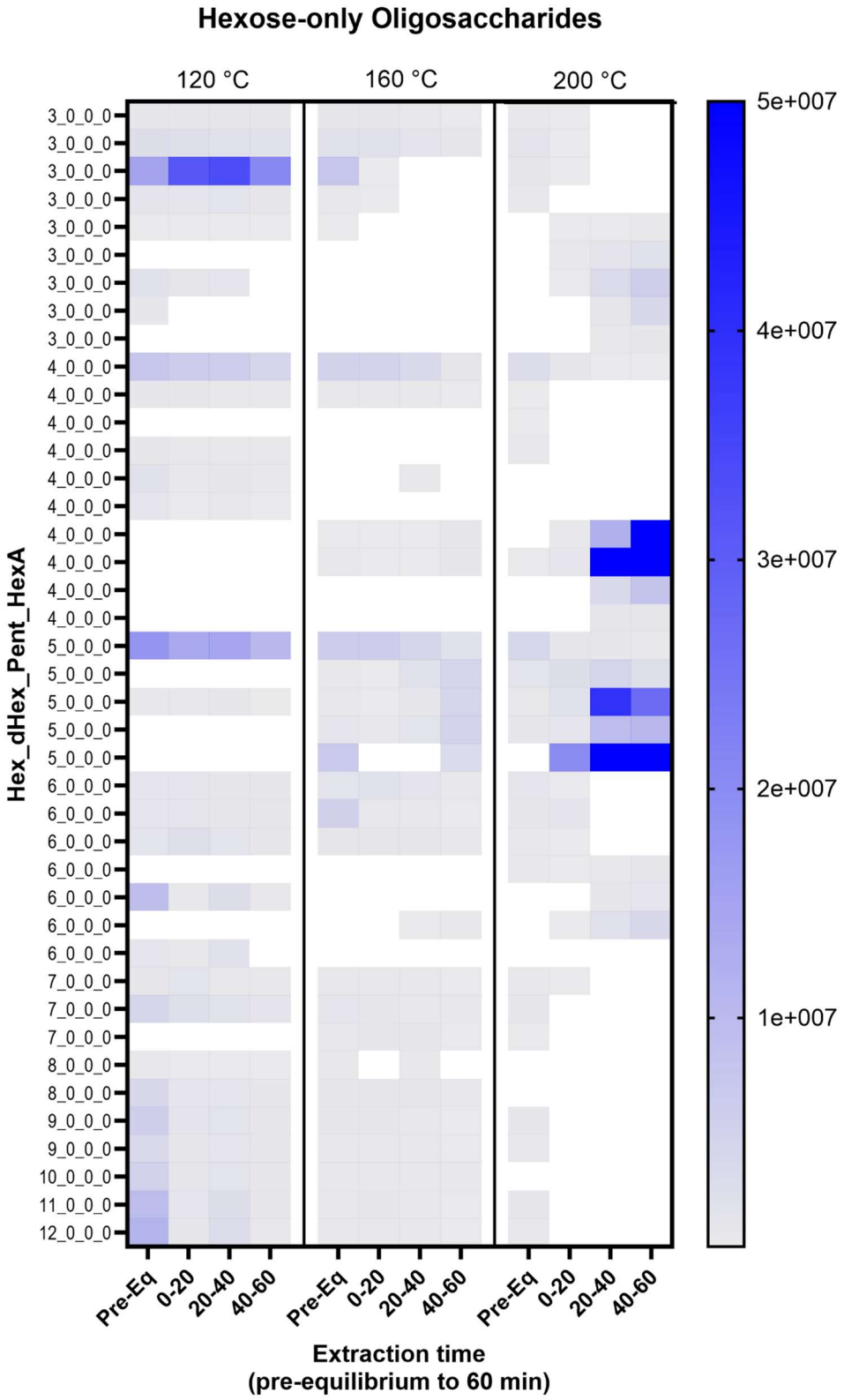
One strategy to organize the data is by sorting based on retention time, which can be indicative of the degrees of polymerization of the oligosaccharides, with longer retention times correlating

with higher degrees of polymerization. Figure 3.5A displays all identified oligosaccharides sorted by retention time, with oligosaccharides having the shortest retention time at the top and the longest at the bottom.

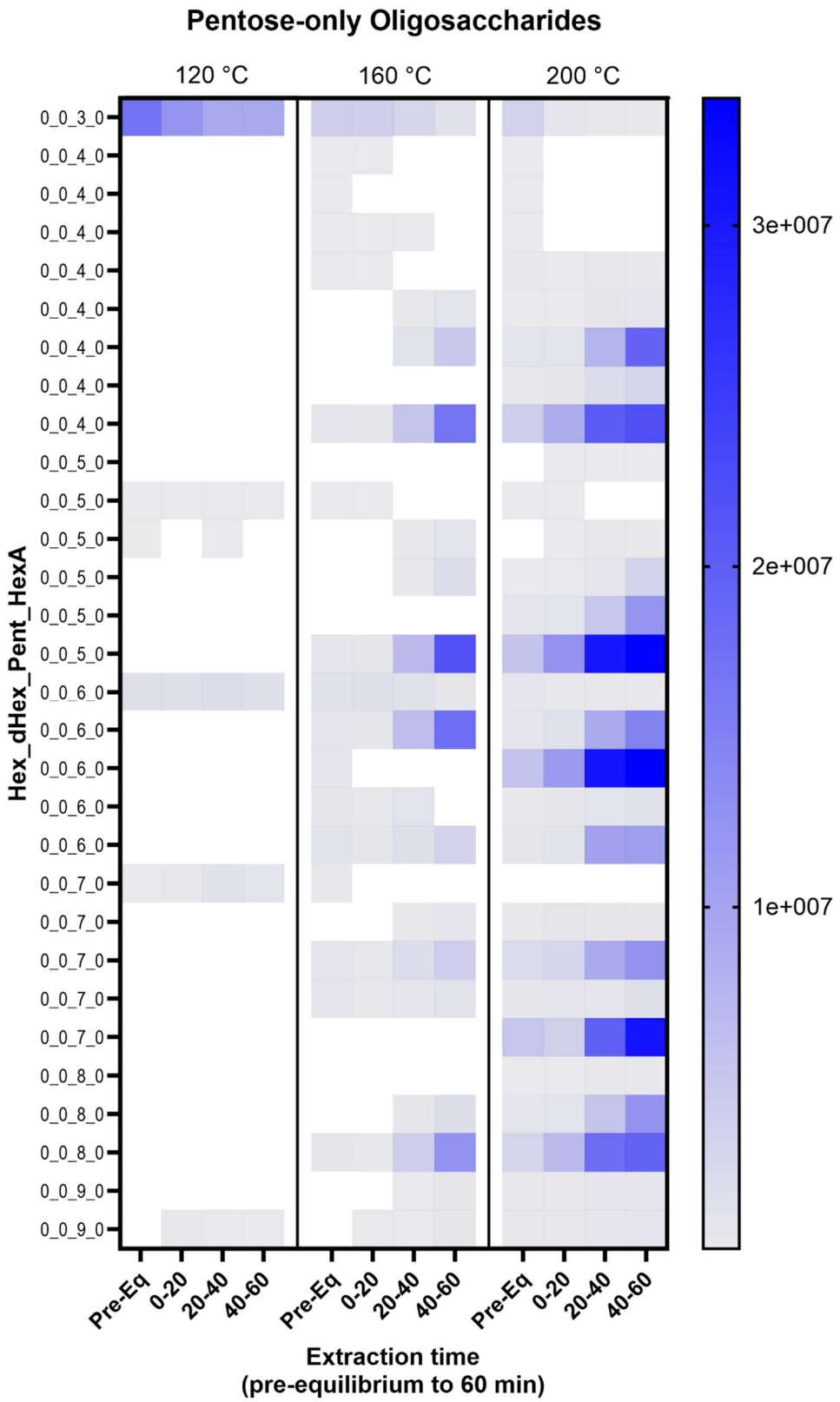
A)


All Oligosaccharides
(Retention Time Sorted)
120 °C
160 °C
200 °C
5e+007
4e+007
3e+007
2e+007
1e+007
Shortest (top) to longest (bottom) retention time
Pre-Eq
0-20
20-40
40-60
Extraction time
(pre-equilibrium to 60 min)

B)


Hexose-only Oligosaccharides
120 °C
160 °C
200 °C
Hex_dHex_Pent_HexA
3_0_0_0
4_0_0_0
5_0_0_0
6_0_0_0
7_0_0_0
8_0_0_0
9_0_0_0
10_0_0_0
11_0_0_0
12_0_0_0
Pre-Eq
0-20
20-40
40-60
Extraction time
(pre-equilibrium to 60 min)
5e+007
4e+007
3e+007
2e+007
1e+007

C)


Pentose-only Oligosaccharides
120 °C
160 °C
200 °C
Hex_dHex_Pent_HexA
0_0_3_0
0_0_4_0
0_0_4_0
0_0_4_0
0_0_4_0
0_0_4_0
0_0_4_0
0_0_4_0
0_0_4_0
0_0_5_0
0_0_5_0
0_0_5_0
0_0_5_0
0_0_5_0
0_0_6_0
0_0_6_0
0_0_6_0
0_0_6_0
0_0_6_0
0_0_7_0
0_0_7_0
0_0_7_0
0_0_7_0
0_0_7_0
0_0_8_0
0_0_8_0
0_0_8_0
0_0_9_0
0_0_9_0
Pre-Eq
0-20
20-40
40-60
Extraction time
(pre-equilibrium to 60 min)
3e+007
2e+007
1e+007

D)

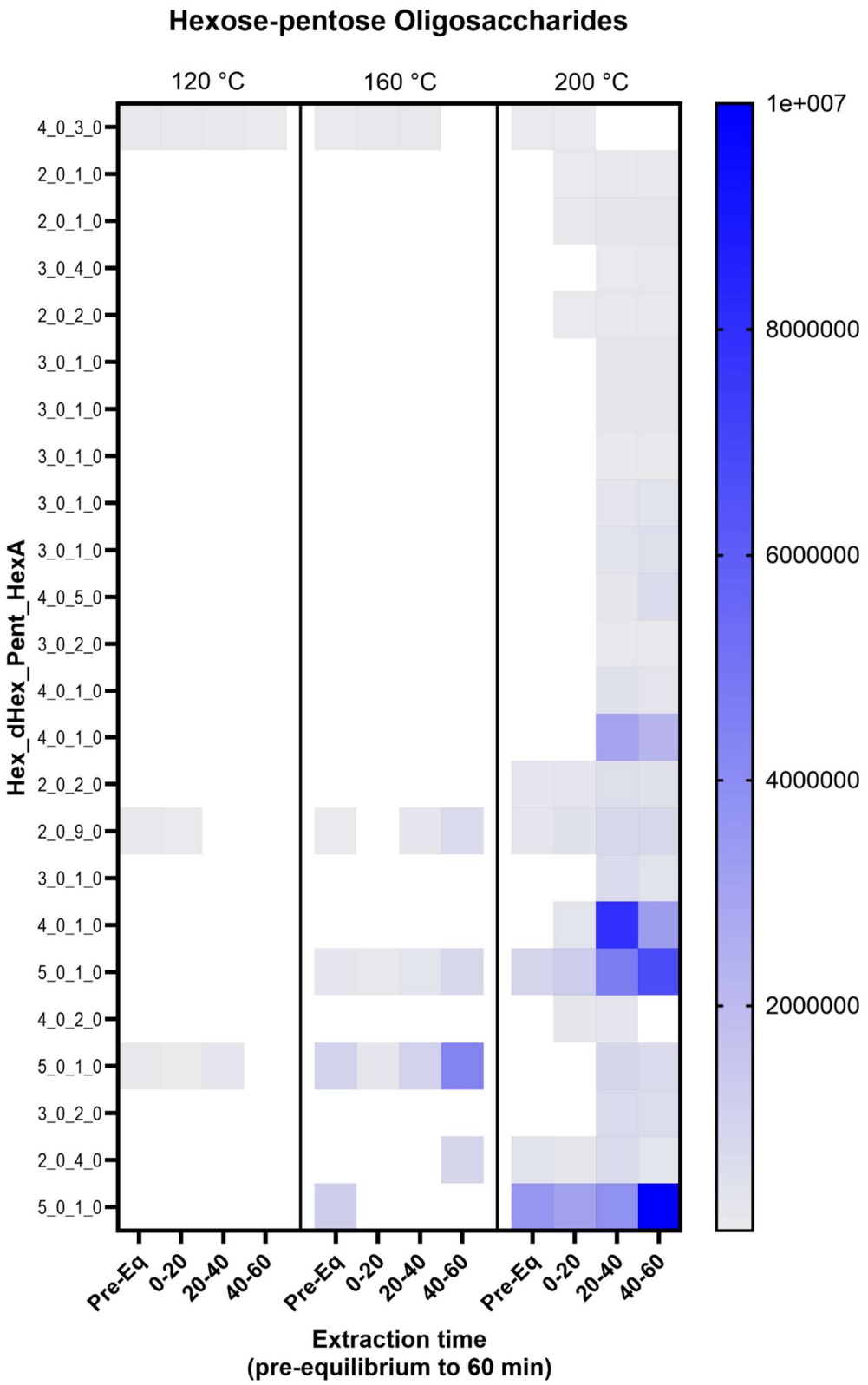

Hexose-pentose Oligosaccharides
120 °C
160 °C
200 °C
Hex_dHex_Pent_HexA
4_0_3_0
2_0_1_0
2_0_1_0
3_0_4_0
2_0_2_0
3_0_1_0
3_0_1_0
3_0_1_0
3_0_1_0
3_0_1_0
4_0_5_0
3_0_2_0
4_0_1_0
4_0_1_0
2_0_2_0
2_0_9_0
3_0_1_0
4_0_1_0
5_0_1_0
4_0_2_0
5_0_1_0
3_0_2_0
2_0_4_0
5_0_1_0
Pre-Eq
0-20
20-40
40-60
Extraction time
(pre-equilibrium to 60 min)
1e+007
8000000
6000000
4000000
2000000

Figure 3.5. Glycoprofile heatmaps of subcritical water extractions of lentil insoluble, with oligosaccharides sorted by retention time (A) and categorized as containing hexose-only (B), pentose-only (C), and hexose-pentose (D). Individual oligosaccharides are reported on each row of the heatmaps, and the four-digit codes along the left y-axis indicate the number of monosaccharides making up the oligosaccharide (Hex_dHex_Pent_HexA). The right color legend indicates peak areas. All samples were analyzed after normalization to soluble carbohydrate content. Oligosaccharides for each subcritical temperature (120 °C, 160 °C, 200 °C) and extraction time (pre-equilibrium, 0 - 20 min, 20 - 40 min, and 40 - 60 min) are reported.

At 200 °C, oligosaccharides with the shorter retention times are predominantly present in the earlier fractions and are not detected with increasing extraction time. Conversely, oligosaccharides with longer retention times are absent in the earlier fractions, but appear in later fractions in high relative abundance. This suggests that oligosaccharides with lower degrees of polymerization are readily extracted at 200 °C, without requiring extended extraction times, while oligosaccharides with higher degrees of polymerization may require longer extraction times at higher temperatures.

In contrast, at the lowest extraction temperature (120 °C), oligosaccharides were observed in high abundance at earlier retention times, but very few oligosaccharides were detected at longer retention times. This suggests that oligosaccharides with lower degrees of polymerization can be extracted at 120 °C, but that this temperature is insufficient to extract the full diversity of oligosaccharides with higher degrees of polymerization as observed at higher temperatures. Approximately 43% of all identified oligosaccharides consisted solely of hexose constituents (hexose-only oligosaccharides) (Fig. 3.3). Among this group, 120 °C consistently yielded the highest number of identified oligosaccharides across all temperatures and extraction times (Fig. 3.4). At 160 °C, the number of identified hexose-only oligosaccharides was slightly lower than that at 120 °C. The highest temperature (200 °C) resulted in the lowest number of hexose-only oligosaccharides, particularly in later fractions under subcritical conditions. At the highest temperature, the pre-equilibrium fraction had the highest abundance. This suggests that hexose-only oligosaccharides are likely to be more completely extracted at the beginning of

extraction under 200 °C but are subsequently degraded at higher temperatures. Further monosaccharide analysis would provide insights into the degradation of these hexose-only oligosaccharides at the highest temperature. These results are also reflected in peak areas as shown in the heatmap of hexose-only oligosaccharides (Figure 3.5B). At the lowest extraction temperature, 120 °C, a variety of hexose-only oligosaccharides are observed fairly consistently among all fractions. On the other hand, at the highest extraction temperature, fewer hexose-only oligosaccharides are observed, with a significant portion of them only appearing in the first or second fraction of extraction.

Approximately 31% of all identified oligosaccharides consisted solely of pentose constituents (pentose-only oligosaccharides) (Fig. 3.3). This group of oligosaccharides was predominantly detected at the highest subcritical water extraction temperature (200 °C), and the number of these oligosaccharides decreased with decreasing subcritical water extraction temperature (Fig. 3.4). For the most part, these pentose-only oligosaccharides are present among all fractions at the highest temperature. At the intermediate temperature (160 °C), an intermediate number of pentose-only oligosaccharides were observed. At this temperature, a few pentose-only oligosaccharides are observed among all fractions, similarly to at 200 °C, however a large portion of these oligosaccharides appear only in later fractions. This indicates that at lower temperatures, longer extraction times are needed to achieve similar results. Finally at the lowest extraction temperature, there was a sharp decrease in the total number of pentose-only oligosaccharides. This may indicate that 120 °C is too low to effectively break down the insoluble matrix. Heat map analysis of the peak area results (Fig. 3C) helps to draw conclusions regarding extraction times. This data reveals that at 200 °C and 160 °C, the relative abundance of these oligosaccharides increases with increasing extraction times. This indicates that for an equivalent amount of carbohydrate, a higher proportion of the extracted oligosaccharides are

pentose-only at longer extraction times (20-40 min and 40-60 min), highlighting the value of subcritical water extraction for extracting this type of oligosaccharides.

Approximately 24% of all identified oligosaccharides were composed of hexose and pentose constituents (hexose-pentose oligosaccharides) (Fig. 3.3). This group of oligosaccharides were predominantly detected under the highest subcritical water extraction temperature (200 °C) (Fig. 3.4). In fact, all hexose-pentose oligosaccharides were identified under the 200 °C extraction conditions. The relative abundance of these oligosaccharides at 200 °C increased with longer extraction times. Most of these oligosaccharides were only detected in the later fractions, further emphasizing the value of extraction under subcritical conditions and the need for sufficiently long extraction times to release these oligosaccharides, even at high temperature. Conversely, out of the 24 oligosaccharides making up the hexose-pentose oligosaccharides group, only 6 and 3 of these oligosaccharides were identified at 160 °C and 120 °C, respectively. This notable decrease in the number of these oligosaccharides at temperatures below 200 °C suggests that there is a minimum temperature threshold required for the release of these oligosaccharides.

Lastly, only 2% of all identified oligosaccharides contained dHex and/or HexA constituents (Fig. 3.3). These oligosaccharides were exclusively detected within subcritical fractions (not pre-equilibrium) at 200 °C, highlighting the value of using subcritical water for the extraction of these oligosaccharides. This profile of this group of oligosaccharides appears similar to that observed for the hexose-pentose oligosaccharides. These findings further underscore the importance of surpassing a temperature threshold to release this type of oligosaccharides.

It is important to note that the reported results were normalized to the total soluble carbohydrate content, which was done to facilitate the generation of a novel oligosaccharide library. In reality, fractions collected at the beginning (pre-equilibrium) and end (40-60 min) of extraction were

orders of magnitude different in total soluble carbohydrate content. This wide range of carbohydrate concentration depending on extraction time was similarly observed by Wiboonsirikul et al. (2013) which showed that in subcritical water extraction of okara, heating at any treatment temperatures for longer than 5 min resulted in decreased carbohydrate content. To provide a more accurate representation of the oligosaccharides obtained through subcritical water extraction, peak abundances were reported without normalization. The unnormalized data is presented in the supplementary figures as unnormalized heatmaps (Supplementary Figures S3.1). The unnormalized data reveals that if an oligosaccharides was detected in the pre-equilibrium fraction, this fraction consistently exhibited the highest abundance.

Discussion

Lentil proteins (AEP, EAEP, Osborne fractionation)

The results of this study found that the albumin-rich fraction of lentils exhibited the highest protein extractability, despite globulins being generally recognized as the major protein class in lentils (making up 42- 48% of total protein) (Bhatty et al., 1976; Neves and Lourenço, 1995). This discrepancy may be attributed to the presence of natural ions in the lentil flour, which effectively form a weak salt solution in the extraction slurry (Bhatty et al., 1976), allowing for the solubilization of salt-soluble globulins by direct extraction with water. Therefore, the albumin-rich fraction likely contains both albumins and globulins, resulting in a higher distribution of proteins extracted solely with water. Similar findings regarding protein distribution were reported by El-Nahry et al (1980) where a significant portion of proteins (42-50%) was extracted in the water-soluble fraction, and a smaller portion (28-35%) was extracted in 10% NaCl. Lastly, the prolamin fraction exhibited a low protein distribution (3%), aligning with previous findings by Bhatty et al. (1976).

Comparing aqueous extraction with enzyme-assisted extraction (EAEP), the use of alkaline protease in the EAEP significantly improved protein extraction yields. Similar increases have been observed in the enzymatic extraction of other pulse proteins, such as chickpea and common bean, using FoodPro® Alkaline Protease (Machida et al., 2022; Yang et al., 2023). The increased protein content in EAEP can be attributed to the increased solubility of smaller protein subunits and peptides generated through proteolysis compared to the intact proteins, thus facilitating their extraction into the aqueous phase (Campbell and Glatz, 2009; de Moura et al., 2008; Souza Almeida et al., 2021; Souza et al., 2019).

Proteomic analysis aligned with the results of total protein extractability, revealing that the prolamin fraction contained the fewest identified proteins (6), while the albumin fraction had the most (32). The albumin and globulin fractions exhibited considerable overlap in terms of protein species, which were also identified in the AEP protein extract. This is expected as all of these extractions were conducted under aqueous conditions, with only slight differences in pH and salt content. The EAEP protein extract had fewer identified proteins compared to AEP (17 vs. 25, respectively), indicating that proteolysis degraded some lentil proteins before proteomics analysis. This was supported by the SDS-PAGE results, which showed minimal intact proteins in EAEP. Notably, the Kunitz-type protein inhibitor, a known antinutritional factor, was not detected in the EAEP. Antinutritional factors can negatively affect the utilization and digestion of plant foods, leading to impaired gastrointestinal functions and metabolic performance (Fekadu Gemede, 2014). These findings suggest that enzymatic extraction of lentil proteins could be an effective strategy for reducing antinutritional factors in the resulting extracts.

This study encountered a prevalent challenge typical in bottom-up proteomics involving variations in the quality of protein databases associated with the target plant species. Since protein identification heavily relies on database searching, the quality of proteomics data

generated is greatly influenced by the protein database's quality. This challenge was particularly evident in the analysis of lentils proteomics, as the available protein database entries were significantly fewer compared to other well-studied legume species (e.g 15,149 entries for lentils, compared to 1,249,705 entries for black beans and 1,412,686 entries for soybeans) (National Center for Biotechnology Information). The limited availability of specific database entries for lentils hindered the acquisition of valuable proteomics data. To overcome this challenge, a cross-species search was conducted. Pea, chickpea, and fava bean showed homology to lentils and were included in the cross-species search. Although generating proteomics data for understudied species poses difficulties, this cross-species search approach offers a viable solution to address this common issue encountered in bottom-up proteomics.

This study aimed to investigate the impact of different extraction methods on the extraction of protein classes in black beans. Sequentially-extracted protein fractions (albumin-, globulin-, prolamin-, and glutelin-rich) were analyzed along with aqueous extraction process and enzyme-assisted aqueous extraction process. SDS-PAGE and proteomics analysis revealed minor variations in protein compositions/profiles among Osborne fractions, with a significant overlap of proteins across different fractions. Sequential fractionation proved useful in understanding the extraction of common bean proteins, despite not providing distinct separation of protein classes. The use of an alkaline protease significantly increased protein extractability compared to aqueous extraction alone and showed potential for reducing antinutritional factors, highlighting the value of using proteomics to identify specific proteins that are extracted by Osborne fractionation or proteins that are susceptible to proteolysis by the alkaline protease. The application of proteomics in this work helps to guide the development of these green extraction technologies, offering detailed insights on the effects of various extraction methods.

Lentil insoluble subcritical water extraction glycoprofiling

Glycomics analysis provided valuable insights into the influence of subcritical water extraction temperature and time on the glycoprofile of insoluble lentil byproducts. The categorization of oligosaccharides based on their constituent monosaccharides (hexose-only, pentose-only, hexose-pentose, others containing dHex and/or HexA) revealed distinct glycoprofiles at different extraction temperatures. At the lowest temperature of 120 °C, the identified oligosaccharides primarily consisted of hexose-only oligosaccharides. This may be attributed to the small amount of starch remaining in the insoluble byproduct following destarching with commercial amylase (after which approximately 20% of starch remained). A very small number of non-hexose-only oligosaccharides were detected at this lowest temperature, suggesting that there is a minimum temperature threshold that is necessary for adequate release of oligosaccharides with a more diverse composition from the insoluble matrix. At the intermediate temperature of 160 °C, the majority of extracted oligosaccharides were hexose-only, with a moderate number of pentose-only oligosaccharides extracted as well. This may indicate the preliminary breakdown of polysaccharides in the insoluble byproduct which likely include cellulose, the most abundant polysaccharides composed to glucose constituents, and hemicellulose, the second most abundant polysaccharide composed of hexose and pentose constituents (Buckeridge et al., 2000). The highest temperature of 200 °C resulted in a relatively smaller number of hexose-only oligosaccharides. However, a significant proportion of oligosaccharides containing pentose constituents (pentose-only oligosaccharides and hexose-pentose oligosaccharides) were attributed only to the highest temperature. This indicates that 200 °C is most effective for breaking down polysaccharides in the insoluble matrix with minimal extraction time. Additionally, the highest temperature yielded a higher number of identified oligosaccharides overall, consistent with findings from other studies. (Chiou et al., 2011; Kataoka et al., 2008; Wiboonsirikul et al., 2007).

The observed variety of glycoprofiles resulting from subcritical extraction at different temperatures demonstrates opportunities for optimization based on the desired oligosaccharide types. For extracting a diversity of oligosaccharides consisting solely of hexoses, 120 °C is a suitable temperature. On the other hand, for extracting a larger diversity of oligosaccharides consisting of hexoses, pentoses, and a combination of hexoses and pentoses, 200 °C offers clear benefits. However, it should be noted that for hexose-only oligosaccharides in particular, increased temperature consistently resulted in a decreased number and abundance of these oligosaccharides. This may be attributed to degradation of hexose-only oligosaccharides into monosaccharides at elevated temperatures. To confirm this, further investigations quantifying monosaccharides in these extracts can provide valuable insights.

It is important to consider the normalization of results when interpreting the findings. Normalization across fractions by carbohydrate content ensured that comparable amounts of carbohydrates were used for LC-MS/MS analysis and maximized the number of detected oligosaccharides for inclusion in the library. However, it should be noted that the actual concentrations of carbohydrates in the samples varied greatly across different fractions. The earlier fractions, particularly the pre-equilibrium fraction, contained much higher carbohydrate content, requiring dilution for analysis. Conversely, the later fractions were very dilute and required extensive freeze drying and centrifugal evaporation to adequately concentrate the fractions for analysis. While this normalization approach is useful for generating novel oligosaccharide libraries and maximizing the number of identified oligosaccharides generated from subcritical water extraction it may not be fully representative of the extraction itself or any potential industrial applications. To obtain more representative results for glycoprofiling studies, analyzing extracts without additional manipulation may provide a truer representation of the extraction process and potentially offer insights that align better with industrial applications.

The diversity of oligosaccharides released during subcritical water extraction, particularly in the later fractions obtained at 200 °C, is promising. However, these results may not be not fully representative of the extracts as they are generated by subcritical water extraction. Analysis of unnormalized data using heatmaps reveals that if an oligosaccharide was detected in the pre-equilibrium fraction, this fraction consistently exhibited the highest abundance. Generally, pre-equilibrium fractions consistently resulted in the most abundant and diverse glycoprofiles compared to later fractions. Therefore, it may be advantageous to focus on shorter extraction times, which still yield desired oligosaccharides without requiring extensive concentration steps and prolonged extraction times. This suggests that achieving a high temperature is of primary importance for releasing oligosaccharides, and that extended extraction times may not be necessary considering the low yields in later fractions. This observation is crucial when taking into account the practical application of this research in industry. However, prior to considering the reduction of extraction time, further studies, such as in-vitro prebiotic studies, should be conducted to determine the specific oligosaccharides profiles that are most desirable, as it is possible that oligosaccharides of interest may only be released with longer extraction times.

This work demonstrated how glycomics techniques can provide valuable insights into the effects of time and temperature during subcritical water extraction. The observed unique profiles of oligosaccharides, depending on the extraction temperature, suggest opportunities for optimization based on oligosaccharides of interest. It is possible to adjust extraction conditions to preferentially extract oligosaccharides of interest, taking into account their structure and composition. The construction of a novel oligosaccharide library for evaluating the breakdown of insoluble lentil byproducts using subcritical water extraction demonstrated the value of this extraction method and the utility of glycomics techniques in guiding the development of these green extraction methods.

Conclusion

The use of proteomics to investigate the green extraction of lentil proteins provided valuable insights into the extraction of protein classes using different extraction methods. The employment of alkaline protease in enzyme-assisted extraction significantly enhanced protein extraction yields, possibly by increasing the solubility of smaller protein subunits and peptides resulting from proteolysis. A cross-species database search was employed to overcome the challenge of a limited protein database specific to lentils. Proteomic analysis confirmed the presence of various protein species in the different protein fractions, and showed considerable overlap between the albumin and globulin fractions. Additionally, proteomics' primary benefit is identification of proteins that stem from various extraction methods. This benefit was demonstrated as proteomics data showed the absence of the Kunitz-type protein inhibitor, an antinutritional factor, in the enzyme-assisted extraction, suggesting the potential for reducing such factors through enzymatic extraction. Overall, the application of proteomics in this study shed light on the effects of different extraction methods on protein composition, providing valuable information for optimizing protein extraction processes and aiding in the development of green extraction technologies.

Glycomics proved to be a valuable tool for investigating the impact of subcritical water extraction temperature and time on the glycoprofile of insoluble lentil byproducts. The categorization of oligosaccharides based on their constituent monosaccharides revealed distinct glycoprofiles at different extraction temperatures. The findings indicated that different temperatures are effective for extracting specific types of oligosaccharides, with 120 °C being suitable for hexose-only oligosaccharides and 200 °C offering clear benefits for a larger diversity of oligosaccharides consisting of both hexoses and pentoses. These results demonstrate opportunities for optimizing extraction conditions based on desired oligosaccharide types. However, it is important to consider the normalization of results when interpreting these findings. While

normalization ensured that a maximum number of oligosaccharides would be included in the lentil insoluble polysaccharide library, the actual concentrations of carbohydrates diminished greatly with increased extraction time. Notably, the abundance and diversity of oligosaccharides released in the pre-equilibrium fraction suggests that shorter extraction times and higher temperatures may be advantageous. Analyzing extracts without additional manipulation may provide a truer representation of the extraction process and align better with industrial applications. Overall, this study demonstrated the value of using glycomics techniques to provide valuable insights into the effects of time and temperature during subcritical water extraction. By utilizing glycomics techniques and constructing a novel oligosaccharide library, this work contributes to the development of green extraction methods and demonstrates the utility of glycomics in guiding such efforts. Further research, including in-vitro prebiotic studies, can build upon these findings to determine the specific oligosaccharide profiles that are most desirable, ultimately enhancing the practical application of this research in the industry.

Acknowledgements

This research was supported by funding by the USDA Pulse Crop Health Initiative (grant number 58-3060-0-044-0).

Supplementary materials

Supplementary Table S3.1. All lentil proteins identified through proteomics analysis. Protein description, accession, the plant database utilized in cross-species search, average mass, PeaksXPro quality score (-10lgP), posttranslational modifications (PTM), and peak areas are reported. Peak areas are separated into the ALB = albumin-rich, GLO = globulin-rich, PRO = prolamin-rich, and GLU = glutelin-rich fractions, and the aqueous extraction process (AEP), and enzyme-assisted aqueous extraction process (EAEP). Peak areas are reported as the sum of the top three peptides.

Description	Accession	Data base	Avg. Mass (Da)	-10 lgP	PTM	Area ALB	Area GLO	Area PRO	Area GLU	Area AEP	Area EAEP
allergen Len c 1.0101 partial	CAD87730.1	lentil	47826	460	Deamidation (NQ)	5.05E+06	1.44E+07	4.89E+03	3.52E+06	1.27E+07	1.61E+06
allergen Len c 1.0102 partial	CAD87731.1	lentil	47467	450	Deamidation (NQ)	4.44E+06	1.15E+07	4.89E+03	3.30E+06	1.05E+07	1.40E+06
convicilin partial	CAB89812.1	lentil	60128	422	Deamidation (NQ)	1.52E+06	4.14E+06	0.00E+00	8.31E+05	3.42E+06	5.44E+05
Lectin	sp|P02870.2|LEC_LENCU	lentil	30352	273	Deamidation (NQ)	4.03E+05	1.25E+06	0.00E+00	1.76E+05	8.32E+05	5.24E+06
lectin	pir||LNLWBA	lentil	25581	253	Deamidation (NQ)	2.98E+05	9.94E+05	0.00E+00	1.01E+05	6.37E+05	3.34E+06
Bowman-Birk type proteinase inhibitor	sp|Q8W4Y8.2|IBB_LENCU	lentil	12266	222	Carbamidomethylation	2.76E+04	1.49E+03	1.59E+03	0.00E+00	1.18E+04	3.28E+03
Chain A Non-specific lipid-transfer protein 2	pdb|2MAL|A	lentil	9292	212	Carbamidomethylation	0.00E+00	0.00E+00	3.11E+04	0.00E+00	0.00E+00	0.00E+00
vicilin type C partial	CAA88357.1	lentil	10331	202	Deamidation (NQ)	2.04E+05	8.78E+05	0.00E+00	1.29E+05	6.31E+05	9.80E+04

Chain D lentil lectin (beta chain)	pdb|2LAL|D	lentil	5709	201		2.24E+05	6.36E+05	0.00E+00	5.36E+04	4.02E+05	2.08E+06
Non-specific lipid-transfer protein 5	sp|A0AT31.1|NLTP5_LENCU	lentil	11686	192	Carbamidomethylation	0.00E+00	0.00E+00	5.89E+03	0.00E+00	0.00E+00	0.00E+00
insecticidal lentil peptide partial	AHG94969.1	lentil	12660	190	Carbamidomethylation	1.70E+05	2.14E+04	7.95E+05	0.00E+00	2.38E+05	3.77E+05
ribulose 1 5-bisphosphate carboxylase/oxygenase large subunit (chloroplast)	AIL56046.1	lentil	52697	180	Carbamidomethylation; Deamidation (NQ)	1.67E+04	0.00E+00	0.00E+00	0.00E+00	2.92E+03	0.00E+00
elongation factor 1-alpha	AYK27098.1	lentil	49249	177	Carbamidomethylation; Deamidation (NQ); Dioxidation (M)	9.85E+04	9.51E+03	0.00E+00	5.57E+04	3.49E+04	3.44E+03
stachyose synthase 1	ALO17655.1	lentil	95966	173	Carbamidomethylation	4.73E+04	0.00E+00	0.00E+00	0.00E+00	4.48E+03	7.94E+03
defensin 5	ATG83510.1	lentil	8228	154	Carbamidomethylation	0.00E+00	0.00E+00	8.18E+02	0.00E+00	0.00E+00	8.24E+03
ADP-glucose pyrophosphorylase small subunit S1 isoform	ACX48912.1	lentil	56232	96	Carbamidomethylation	1.25E+04	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
chloroplast copper/zinc superoxide dismutase	QEJ74041.1	lentil	20552	87	Carbamidomethylation	1.18E+04	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
ADP-glucose pyrophosphorylase large subunit L1 isoform	ACW82825.1	lentil	56244	87	Carbamidomethylation	5.52E+04	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
Vicilin	sp|P13918.2|VCLC_PEA	pea	52231	312	Deamidation (NQ)	3.69E+06	1.10E+07	2.92E+03	1.07E+06	9.73E+06	1.27E+06
Vicilin partial	CAF25233.1	pea	47298	278	Deamidation (NQ)	2.07E+06	4.98E+06	5.67E+02	1.14E+06	5.44E+06	6.65E+05

Seed linoleate 9S-lipoxygenase-3	sp|P09918.1|LOX3_PEA	pea	97629	257	Carbamidomethylation; Deamidation (NQ)	1.97E+06	8.85E+04	0.00E+00	2.13E+05	8.27E+05	9.31E+04
vicilin 47k	CBK38917.1	pea	49466	256	Deamidation (NQ)	1.81E+06	4.32E+06	0.00E+00	1.14E+06	4.61E+06	7.03E+05
Legumin J	sp|P05692.1|LEGJ_PEA	pea	56895	236	Carbamidomethylation; Deamidation (NQ)	1.65E+05	1.42E+06	0.00E+00	2.47E+04	7.50E+05	2.37E+04
Convicilin	sp|P13915.1|CVCA_PEA	pea	66990	219	Deamidation (NQ)	6.61E+05	2.42E+06	0.00E+00	3.12E+05	1.84E+06	3.77E+05
legumin K - garden pea	pir||S26688	pea	56276	212	Carbamidomethylation; Deamidation (NQ)	6.32E+05	1.59E+06	0.00E+00	9.35E+03	9.16E+05	8.33E+04
heat shock protein hsp70	CAA67867.1	pea	71167	211		2.80E+05	6.80E+03	0.00E+00	0.00E+00	7.15E+04	4.77E+02
cvc partial	CAP06315.1	pea	62086	210	Deamidation (NQ)	8.11E+05	1.91E+06	0.00E+00	2.70E+05	1.50E+06	3.11E+05
Legumin A2	sp|P15838.1|LEGA2_PEA	pea	59270	207	Carbamidomethylation; Deamidation (NQ)	1.62E+06	3.98E+06	0.00E+00	2.07E+06	2.62E+06	1.38E+06
P54 protein	CAA72090.1	pea	54662	205	Deamidation (NQ)	1.82E+05	9.66E+05	0.00E+00	2.57E+06	7.23E+05	6.07E+04
Provicilin	sp|P02855.1|VCLA_PEA	pea	31540	201	Deamidation (NQ)	3.63E+05	4.05E+06	6.16E+02	1.64E+05	3.22E+06	9.36E+04
Legumin A	sp|P02857.1|LEGA_PEA	pea	58805	184	Carbamidomethylation; Deamidation (NQ)	1.05E+06	2.20E+06	0.00E+00	5.14E+05	1.68E+06	1.15E+06
lipoxygenase	CAA55318.1	pea	97002	172		7.87E+04	3.62E+03	0.00E+00	2.87E+03	2.02E+04	4.09E+03
legumin (minor small)	CAA47809.1	pea	64873	171	Carbamidomethylation; Deamidation (NQ)	7.47E+05	1.43E+06	0.00E+00	6.14E+05	1.12E+06	5.87E+03

Seed biotin-containing protein SBP65	sp|Q41060.1|SBP65	pea	59554	171	Deamidation (NQ)	7.52E+04	2.89E+04	2.98E+03	4.67E+04	1.28E+05	0.00E+00
PsHSC71.0	CAA83548.1	pea	71004	152		9.11E+04	0.00E+00	0.00E+00	0.00E+00	2.05E+04	0.00E+00
Alcohol dehydrogenase 1	sp|P12886.1|ADH1_PEA	pea	41155	150	Deamidation (NQ)	2.55E+05	5.54E+02	0.00E+00	1.79E+04	8.61E+04	2.79E+03
RuBisCO large subunit-binding protein subunit alpha chloroplastic	sp|P08926.2|RUBA_PEA	pea	61979	147		9.68E+03	0.00E+00	0.00E+00	0.00E+00	2.73E+03	0.00E+00
heat shock protein	AAN74634.1	pea	18055	144		9.36E+04	4.95E+03	0.00E+00	1.76E+03	7.88E+04	1.80E+03
Glyceraldehyde-3-phosphate dehydrogenase cytosolic	sp|P34922.1|G3PC_PEA	pea	36609	139		1.94E+05	2.81E+03	0.00E+00	8.89E+03	2.67E+03	0.00E+00
Phosphoglucomutase cytoplasmic	sp|Q9SM60.1|PGMC_PEA	pea	63325	135		1.98E+04	0.00E+00	0.00E+00	0.00E+00	4.76E+03	0.00E+00
Histone H4	sp|P62788.2|H4_PEA	pea	11409	131		0.00E+00	1.25E+04	0.00E+00	1.67E+05	2.14E+04	6.80E+03
vitamin B-12-independent methionine synthase	AIE47233.1	pea	84317	129	Carbamidomethylation; Deamidation (NQ)	6.14E+04	0.00E+00	0.00E+00	0.00E+00	9.51E+03	0.00E+00
Fructose-bisphosphate aldolase cytoplasmic isozyme 2	sp|P46257.1|ALF2_PEA	pea	38491	129		4.68E+04	1.78E+03	0.00E+00	0.00E+00	1.77E+04	0.00E+00
17.9 kDa heat shock protein (hsp17.9) partial	AAA33671.1	pea	17644	128	Deamidation (NQ); Dioxidation (M)	9.11E+04	1.14E+04	0.00E+00	0.00E+00	5.43E+04	0.00E+00
F1 ATPase	BAA20135.1	pea	60151	127		2.85E+04	3.20E+03	0.00E+00	1.89E+04	5.20E+03	0.00E+00

alpha-galactosidase 1	CAF34023.1	pea	44964	126	Carbamidomethylation	3.56E+02	1.00E+04	0.00E+00	0.00E+00	1.21E+04	0.00E+00
Nucleoside diphosphate kinase 1	sp|P47922.1|NDK1	pea	16463	125		6.68E+04	4.39E+03	0.00E+00	1.34E+04	2.40E+04	7.19E+04
Polyubiquitin	sp|P69322.2|UBIQP_PEA	pea	42699	118		0.00E+00	0.00E+00	3.28E+02	0.00E+00	5.99E+03	0.00E+00
Albumin-2	sp|P08688.1|ALB2_PEA	pea	26238	108		6.32E+05	4.05E+04	0.00E+00	0.00E+00	2.23E+05	4.26E+04
Ribulose bisphosphate carboxylase large chain	sp|P04717.3|RBL_PEA	pea	52763	107	Carbamidomethylation	1.17E+04	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
Ran1	ABM73376.1	pea	25330	107	Carbamidomethylation	1.80E+04	0.00E+00	0.00E+00	0.00E+00	2.47E+03	1.24E+03
alpha 1 4 glucan-phosphorylase partial	AIS70806.1	pea	7028	98	Carbamidomethylation	5.71E+04	0.00E+00	0.00E+00	0.00E+00	1.07E+04	0.00E+00
DEAD box RNA helicase	AAN74635.1	pea	46881	96		8.51E+03	1.08E+03	0.00E+00	0.00E+00	2.11E+03	0.00E+00
Ferritin-1 chloroplastic	sp|P19975.2|FRI1_PEA	pea	28619	94		2.44E+04	3.23E+03	0.00E+00	0.00E+00	1.88E+03	6.14E+03
Fructose-bisphosphate aldolase 1 chloroplastic	sp|Q01516.1|ALFC1_PEA	pea	38657	92	Carbamidomethylation	6.34E+03	0.00E+00	0.00E+00	0.00E+00	3.19E+03	0.00E+00
Superoxide dismutase [Cu-Zn]	sp|Q02610.2|SODC_PEA	pea	15323	90	Carbamidomethylation	1.80E+04	2.59E+02	0.00E+00	0.00E+00	3.45E+03	0.00E+00
cytosolic phosphoglycerate kinase	AAF85975.1	pea	42287	81		4.76E+04	0.00E+00	0.00E+00	0.00E+00	9.37E+03	0.00E+00
22.7 kDa class IV heat shock protein	sp|P19244.1|HSP41_PEA	pea	22734	81	Deamidation (NQ)	5.22E+03	0.00E+00	0.00E+00	2.71E+04	0.00E+00	1.36E+03

short-chain alcohol dehydrogenase	AAF04193.1	pea	28210	80		4.48E+03	0.00E+00	0.00E+00	0.00E+00	0.00E+00	9.86E+03
ATP synthase subunit alpha mitochondrial	sp|P05493.2|ATPAM_PEA	pea	55045	78		2.07E+04	4.20E+03	0.00E+00	0.00E+00	0.00E+00	0.00E+00
Kunitz-type trypsin inhibitor-like 1 protein	sp|Q41015.2|PIP21_PEA	pea	23792	75	Carbamidomethylation	0.00E+00	0.00E+00	0.00E+00	0.00E+00	3.59E+03	0.00E+00
Superoxide dismutase [Cu-Zn] chloroplastic	sp|P11964.1|SODCP_PEA	pea	20626	73	Carbamidomethylation	1.18E+04	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
Heat shock 70 kDa protein mitochondrial	sp|P37900.1|HSP7M_PEA	pea	72301	72		2.31E+03	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
vicilin-like	XP_004492829.1	chick pea	51087	160	Deamidation (NQ)	1.54E+06	5.64E+06	1.74E+03	8.21E+05	3.94E+06	5.12E+05
lectin	ADO32620.1	chick pea	30280	148	Deamidation (NQ)	1.87E+05	1.20E+06	0.00E+00	0.00E+00	6.10E+05	5.06E+06
heat shock 70 kDa protein	XP_004507753.1	chick pea	71210	141	Carbamidomethylation	2.28E+05	5.99E+03	0.00E+00	0.00E+00	6.88E+04	0.00E+00
heat shock cognate 70 kDa protein 2	XP_027188896.1	chick pea	106304	137		1.43E+05	0.00E+00	0.00E+00	0.00E+00	4.33E+04	0.00E+00
seed linoleate 9S-lipoxygenase-3	XP_004486857.1	chick pea	97493	133		3.77E+05	3.23E+02	0.00E+00	7.68E+04	1.09E+05	0.00E+00
NADPH-dependent aldehyde reductase 1 chloroplastic-like	XP_004494625.2	chick pea	35512	124		2.55E+05	7.22E+03	0.00E+00	4.18E+04	4.17E+04	3.78E+03
provicilin-like	XP_004496703.1	chick pea	64651	123	Deamidation (NQ)	6.45E+05	1.67E+06	0.00E+00	2.75E+05	1.81E+06	3.14E+05

cytosolic malate dehydrogenase	CAC10208.1	chick pea	35498	114	Carbamidomethylation	6.56E+04	0.00E+00	0.00E+00	0.00E+00	1.22E+04	0.00E+00
histone H4	XP_004511025.1	chick pea	11409	113		0.00E+00	1.25E+04	0.00E+00	1.67E+05	2.14E+04	6.80E+03
Chain A legumin-like protein	pdb|5GYL|A	chick pea	54173	113	Carbamidomethylation; Deamidation (NQ)	5.79E+05	1.80E+06	0.00E+00	1.49E+06	9.50E+05	2.34E+05
legumin A-like	XP_004493780.1	chick pea	59343	111	Carbamidomethylation	3.22E+05	1.39E+06	0.00E+00	0.00E+00	1.04E+06	9.18E+04
ATP synthase subunit beta mitochondrial-like	XP_012570234.1	chick pea	60008	110		2.85E+04	3.20E+03	0.00E+00	1.05E+04	5.20E+03	0.00E+00
ruBisCO large subunit-binding protein subunit alpha chloroplastic	XP_004506047.1	chick pea	62057	108		5.96E+03	0.00E+00	0.00E+00	0.00E+00	2.73E+03	0.00E+00
alcohol dehydrogenase 1	XP_004502579.1	chick pea	41069	105		2.53E+05	5.54E+02	0.00E+00	2.92E+03	8.55E+04	2.79E+03
fructose-bisphosphate aldolase 1 chloroplastic	XP_004507508.1	chick pea	43129	105	Carbamidomethylation	9.86E+03	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
malate dehydrogenase	XP_004495378.1	chick pea	35726	103	Carbamidomethylation	5.12E+04	0.00E+00	0.00E+00	0.00E+00	6.91E+03	0.00E+00
legumin J	XP_012569358.1	chick pea	62715	101	Carbamidomethylation	1.08E+05	1.20E+05	0.00E+00	5.06E+04	1.59E+05	5.32E+03
luminal-binding protein	XP_004503579.1	chick pea	73558	100		5.51E+04	0.00E+00	0.00E+00	8.88E+03	8.98E+03	0.00E+00
seed linoleate 9S-lipoxygenase-2	XP_027189582.1	chick pea	96937	99		1.84E+05	1.16E+04	0.00E+00	0.00E+00	5.05E+04	0.00E+00

1-Cys peroxiredoxin isoform X2	XP_004504653.1	chick pea	24364	99	Deamidation (NQ)	3.14E+05	0.00E+00	0.00E+00	0.00E+00	1.88E+05	0.00E+00
alcohol dehydrogenase 1-like	XP_004502578.1	chick pea	41021	95		3.01E+05	5.54E+02	0.00E+00	2.92E+03	8.55E+04	2.79E+03
albumin-2-like	NP_001351664.1	chick pea	26148	93		5.39E+05	6.58E+04	0.00E+00	0.00E+00	2.38E+05	0.00E+00
18.5 kDa class I heat shock protein	XP_004505083.1	chick pea	18384	89	Dioxidation (M)	5.71E+04	0.00E+00	0.00E+00	0.00E+00	4.88E+04	0.00E+00
heat shock protein 83	XP_004516872.1	chick pea	80331	87		1.05E+04	0.00E+00	0.00E+00	0.00E+00	1.60E+04	0.00E+00
leucine aminopeptidase 1-like	XP_004507738.1	chick pea	59867	86		2.57E+04	0.00E+00	0.00E+00	0.00E+00	1.68E+05	0.00E+00
glyceraldehyde-3-phosphate dehydrogenase cytosolic	XP_004502328.1	chick pea	37021	86		1.50E+05	8.62E+02	0.00E+00	8.89E+03	1.59E+04	0.00E+00
fructose-bisphosphate aldolase 6 cytosolic isoform X2	XP_004497605.1	chick pea	38351	86		2.49E+04	1.78E+03	0.00E+00	0.00E+00	2.03E+03	0.00E+00
ribulose-1 5-bisphosphate carboxylase/oxygenase large subunit (chloroplast)	ACH41054.1	chick pea	52686	85	Carbamidomethylation	1.17E+04	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
17.5 kDa class I heat shock protein-like	XP_004490930.1	chick pea	17443	85	Dioxidation (M)	6.84E+04	0.00E+00	0.00E+00	0.00E+00	4.88E+04	0.00E+00
nucleoside diphosphate kinase 1	XP_004515334.1	chick pea	16486	84		2.97E+04	1.74E+03	0.00E+00	1.34E+04	2.40E+04	7.19E+04
uncharacterized protein LOC101498325	XP_027190363.1	chick pea	8102	82		9.86E+03	0.00E+00	0.00E+00	0.00E+00	3.26E+03	0.00E+00

endoplasmin homolog isoform X2	XP_004499283.1	chick pea	93747	81		2.77E+04	1.10E+03	0.00E+00	0.00E+00	0.00E+00	0.00E+00
60S ribosomal protein L8-3-like	XP_004492717.1	chick pea	28101	81		2.44E+04	2.48E+03	0.00E+00	0.00E+00	3.30E+03	0.00E+00
eukaryotic initiation factor 4A	NP_001352094.1	chick pea	46884	80		1.73E+04	0.00E+00	0.00E+00	0.00E+00	2.11E+03	0.00E+00
non-specific lipid-transfer protein precursor	NP_001296611.2	chick pea	11588	79	Carbamidomethylation; Deamidation (NQ)	0.00E+00	0.00E+00	5.46E+03	0.00E+00	0.00E+00	0.00E+00
ubiquitin-40S ribosomal protein S27a	XP_004489824.1	chick pea	17698	79		0.00E+00	0.00E+00	3.28E+02	0.00E+00	5.99E+03	0.00E+00
peptidyl-prolyl cis-trans isomerase	XP_004500742.1	chick pea	20709	78		3.62E+04	0.00E+00	0.00E+00	3.42E+04	4.27E+03	0.00E+00
alpha-1 4 glucan phosphorylase L isozyme chloroplastic/amyloplastic	XP_004489452.1	chick pea	111311	77		2.20E+04	0.00E+00	0.00E+00	0.00E+00	7.98E+03	0.00E+00
UTP--glucose-1-phosphate uridylyltransferase	XP_004490785.1	chick pea	51565	76		5.86E+03	0.00E+00	0.00E+00	0.00E+00	3.70E+03	0.00E+00
phosphoglucomutase cytoplasmic	XP_004503379.1	chick pea	63729	75		1.53E+03	0.00E+00	0.00E+00	0.00E+00	2.83E+02	0.00E+00
annexin-like protein RJ4	XP_027186388.1	chick pea	36262	75		0.00E+00	0.00E+00	0.00E+00	1.38E+04	9.16E+03	0.00E+00
40S ribosomal protein S3-3	XP_004497806.1	chick pea	26436	75	Carbamidomethylation	1.08E+04	0.00E+00	0.00E+00	0.00E+00	8.70E+03	5.63E+03
heat shock cognate protein 80	XP_004500151.1	chick pea	80038	73		1.35E+04	0.00E+00	0.00E+00	0.00E+00	5.88E+03	0.00E+00

ruBisCO large subunit-binding protein subunit beta chloroplastic	XP_012567814.1	chick pea	62977	73		2.32E+03	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
peroxygenase isoform X1	XP_004495300.1	chick pea	27135	71	Carbamidomethylation	4.04E+04	0.00E+00	0.00E+00	5.30E+04	0.00E+00	0.00E+00
elongation factor 1-alpha	NP_001352092.1	chick pea	49330	70	Carbamidomethylation	0.00E+00	0.00E+00	0.00E+00	3.28E+03	0.00E+00	3.44E+03
triosephosphate isomerase cytosolic	XP_004487007.1	chick pea	27102	68		6.84E+03	0.00E+00	0.00E+00	0.00E+00	2.51E+04	0.00E+00
probable phospholipid hydroperoxide glutathione peroxidase	XP_004503771.1	chick pea	26618	68		2.53E+04	1.88E+03	0.00E+00	0.00E+00	9.91E+03	0.00E+00
60S ribosomal protein L6-3-like	XP_004504017.1	chick pea	26191	65		2.81E+03	0.00E+00	0.00E+00	0.00E+00	3.44E+03	7.31E+03
40S ribosomal protein S4-1-like	XP_004491013.1	chick pea	29997	65	Carbamidomethylation	7.68E+03	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
sucrose-binding protein-like	XP_004495184.1	chick pea	53803	62		4.41E+03	0.00E+00	0.00E+00	0.00E+00	4.21E+03	0.00E+00
legumin A2 primary translation product	CAA38758.1	fava bean	56680	366	Carbamidomethylation; Deamidation (NQ)	2.04E+06	4.51E+06	0.00E+00	1.52E+06	3.48E+06	3.79E+06
Alpha-1 4 glucan phosphorylase L isozyme chloroplastic/amyloplastic	sp|P53536.2|PHSL_VICFA	fava bean	113580	321	Carbamidomethylation	2.70E+05	7.77E+03	0.00E+00	4.32E+04	7.34E+04	0.00E+00
convicilin partial	CAP06335.1	fava bean	57501	263	Deamidation (NQ)	7.18E+05	1.91E+06	0.00E+00	6.89E+05	1.50E+06	3.11E+05
storage protein	CAA32455.1	fava bean	34234	221	Carbamidomethylation; Deamidation (NQ)	1.23E+06	2.09E+06	0.00E+00	6.81E+05	1.72E+06	7.73E+04

ribulose-1 5-bisphosphate carboxylase/oxygenase large subunit partial (chloroplast)	AWM67354.1	fava bean	52320	166	Carbamidomethylation	1.17E+04	0.00E+00	0.00E+00	0.00E+00	2.48E+03	0.00E+00
legumin	CAA81262.1	fava bean	64502	163	Carbamidomethylation; Deamidation (NQ)	1.19E+05	2.44E+05	0.00E+00	0.00E+00	2.02E+05	0.00E+00
Elongation factor 1-alpha	sp|O24534.1|EF1A_VICFA	fava bean	49244	145	Carbamidomethylation; Deamidation (NQ)	8.06E+04	0.00E+00	0.00E+00	4.35E+03	0.00E+00	3.44E+03
GTP-binding nuclear protein Ran/TC4	sp|P38548.1|RAN_VICFA	fava bean	25290	137	Carbamidomethylation	1.80E+04	0.00E+00	0.00E+00	0.00E+00	2.47E+03	1.24E+03
polyubiquitin	CAA10056.1	fava bean	25685	130		0.00E+00	0.00E+00	1.27E+02	0.00E+00	5.99E+03	0.00E+00
putative sucrose binding protein	CAC27161.1	fava bean	54614	120		0.00E+00	5.45E+05	0.00E+00	1.27E+06	2.23E+05	0.00E+00
Probable sucrose-phosphate synthase	sp|Q43876.1|SPSA_VICFA	fava bean	118204	107		1.83E+03	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
lipoxygenase	CAA97845.1	fava bean	96504	105		1.04E+04	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
14-3-3-like protein A	sp|P42653.1|1433A_VICFA	fava bean	29420	96		0.00E+00	0.00E+00	0.00E+00	5.00E+03	0.00E+00	0.00E+00
Alpha-glucan phosphorylase H isozyme	sp|P53537.1|PHSH_VICFA	fava bean	95924	90		5.56E+04	4.11E+03	0.00E+00	0.00E+00	0.00E+00	0.00E+00
Glucose-1-phosphate adenylyltransferase small subunit 2 chloroplastic	sp|P52417.1|GLGS2_VICFA	fava bean	56060	90	Carbamidomethylation	1.25E+04	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00

Supplementary Table S3.2. Compositions, neutral mass, and retention time (minutes) of all oligosaccharides confirmed by tandem MS/MS in at least one of the lentil subcritical water extractions.

Hex_dHex_Pent_HexA	Mass	Retention time (min)	Hex	dHex	Pent	HexA
2_0_1_0	474.159	12.0	2	0	1	0
2_0_1_0	474.159	12.8	2	0	1	0
3_0_0_0	504.169	3.2	3	0	0	0
3_0_0_0	504.169	5.5	3	0	0	0
3_0_0_0	504.169	6.4	3	0	0	0
3_0_0_0	504.169	8.5	3	0	0	0
3_0_0_0	504.169	10.1	3	0	0	0
3_0_0_0	504.169	11.9	3	0	0	0
3_0_0_0	504.169	16.5	3	0	0	0
3_0_0_0	504.169	17.3	3	0	0	0
3_0_0_0	504.169	19.5	3	0	0	0
3_0_0_0	504.169	22.1	3	0	0	0
3_0_0_0	504.169	22.6	3	0	0	0
3_0_0_0	504.169	23.7	3	0	0	0
3_0_0_0	504.169	24.8	3	0	0	0
0_0_4_0	546.180	12.7	0	0	4	0

0_0_4_0	546.180	14.5	0	0	4	0
0_0_4_0	546.180	17.4	0	0	4	0
0_0_4_0	546.180	20.3	0	0	4	0
0_0_4_0	546.180	22.7	0	0	4	0
0_0_4_0	546.180	24.9	0	0	4	0
0_0_4_0	546.180	26.5	0	0	4	0
2_0_2_0	606.201	15.3	2	0	2	0
2_0_2_0	606.201	24.4	2	0	2	0
3_0_1_0	636.211	18.9	3	0	1	0
3_0_1_0	636.211	19.6	3	0	1	0
3_0_1_0	636.211	19.7	3	0	1	0
3_0_1_0	636.211	21.8	3	0	1	0
3_0_1_0	636.211	23.6	3	0	1	0
4_0_0_0	666.222	6.4	4	0	0	0
4_0_0_0	666.222	12.5	4	0	0	0
4_0_0_0	666.222	14.1	4	0	0	0
4_0_0_0	666.222	19.7	4	0	0	0
4_0_0_0	666.222	23.5	4	0	0	0

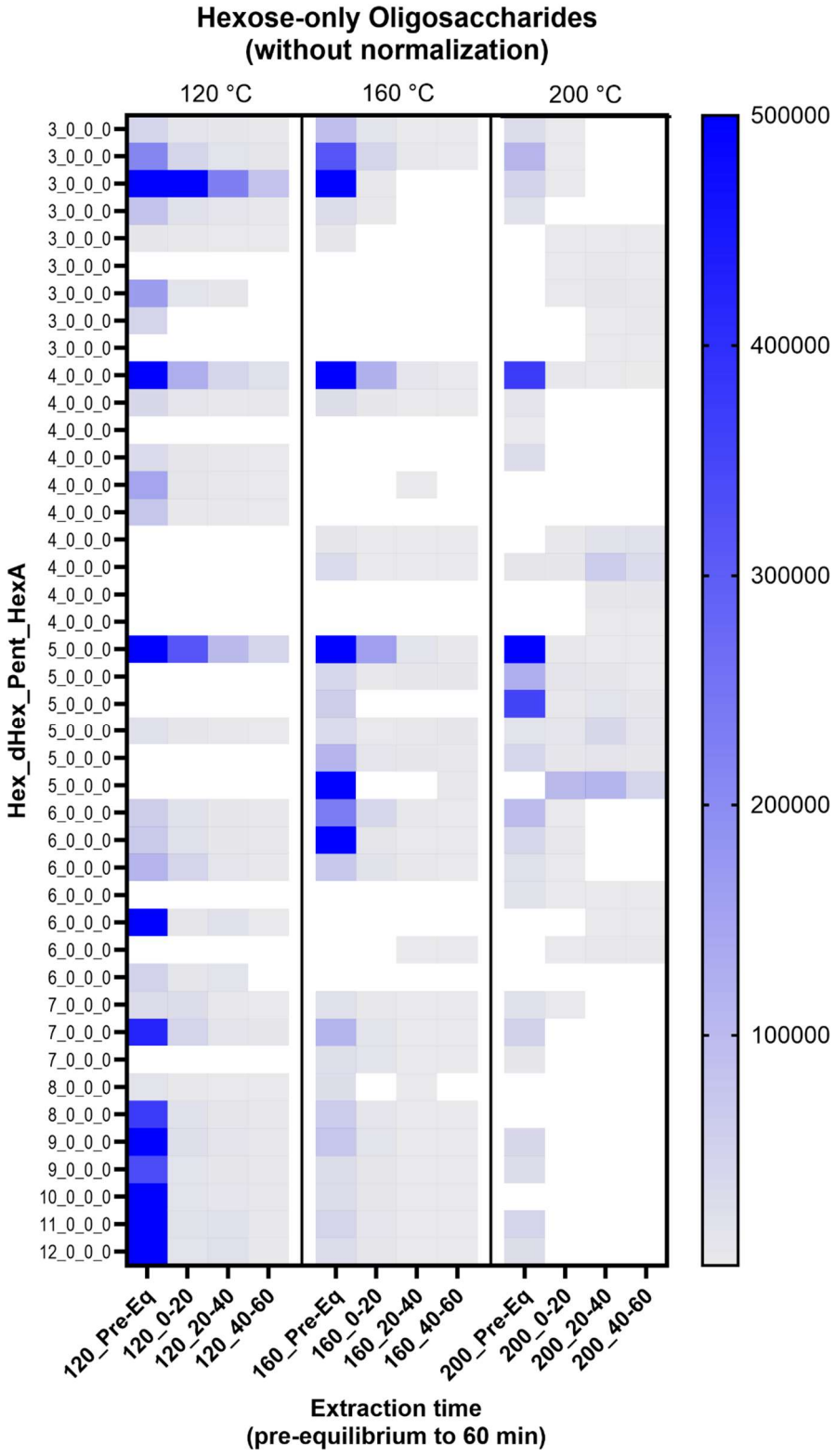
4_0_0_0	666.222	24.7	4	0	0	0
4_0_0_0	666.222	27.5	4	0	0	0
4_0_0_0	666.222	9.4	4	0	0	0
4_0_0_0	666.222	25.6	4	0	0	0
0_0_5_0	678.222	16.4	0	0	5	0
0_0_5_0	678.222	19.2	0	0	5	0
0_0_5_0	678.222	23.1	0	0	5	0
0_0_5_0	678.222	25.1	0	0	5	0
0_0_5_0	678.222	26.4	0	0	5	0
0_0_5_0	678.222	28.6	0	0	5	0
0_0_5_0	678.222	30.4	0	0	5	0
0_0_5_0	678.222	32.4	0	0	5	0
3_0_2_0	768.254	29.0	3	0	2	0
4_0_1_0	798.264	19.7	4	0	1	0
4_0_1_0	798.264	22.2	4	0	1	0
4_0_1_0	798.264	25.3	4	0	1	0
4_0_1_0	798.264	27.3	4	0	1	0
4_0_1_0	798.264	28.6	4	0	1	0

0_0_6_0	810.264	20.7	0	0	6	0
0_0_6_0	810.264	23.1	0	0	6	0
0_0_6_0	810.264	24.6	0	0	6	0
0_0_6_0	810.264	26.3	0	0	6	0
0_0_6_0	810.264	28.2	0	0	6	0
0_0_6_0	810.264	29.5	0	0	6	0
0_0_6_0	810.264	32.5	0	0	6	0
0_0_6_0	810.264	10.6	0	0	6	0
5_0_0_0	828.275	8.4	5	0	0	0
5_0_0_0	828.275	10.5	5	0	0	0
5_0_0_0	828.275	12.5	5	0	0	0
5_0_0_0	828.275	24.4	5	0	0	0
5_0_0_0	828.275	28.5	5	0	0	0
5_0_0_0	828.275	30.5	5	0	0	0
5_0_0_0	828.275	32.2	5	0	0	0
4_0_2_0	930.306	27.7	4	0	2	0
0_0_7_0	942.307	18.2	0	0	7	0
0_0_7_0	942.307	25.5	0	0	7	0

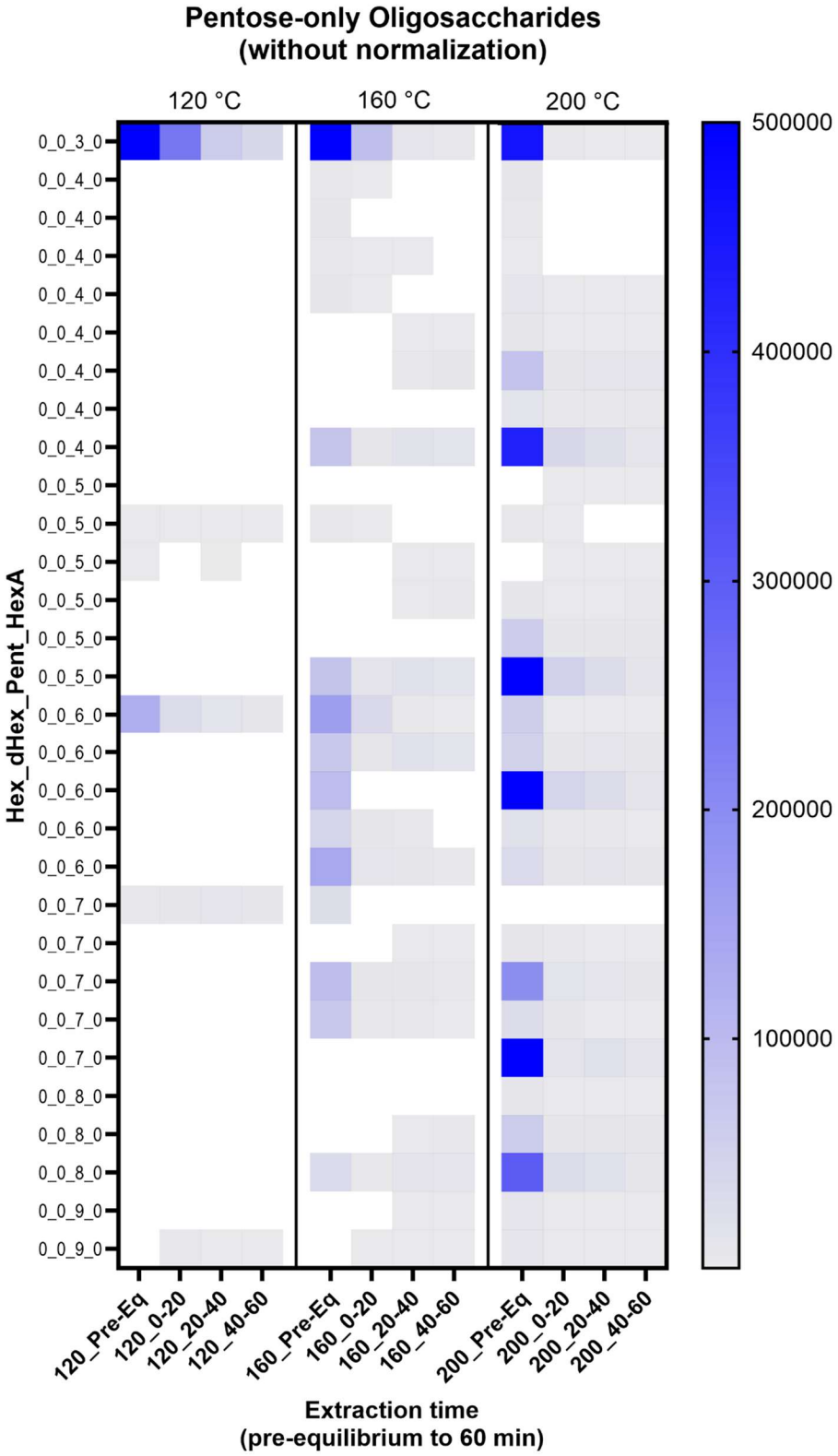
0_0_7_0	942.307	29.0	0	0	7	0
0_0_7_0	942.307	29.5	0	0	7	0
0_0_7_0	942.307	30.6	0	0	7	0
0_0_7_0	942.307	32.6	0	0	7	0
5_0_1_0	960.317	29.2	5	0	1	0
5_0_1_0	960.317	30.8	5	0	1	0
5_0_1_0	960.317	31.7	5	0	1	0
3_2_0_1	972.317	15.3	3	2	0	1
6_0_0_0	990.327	10.6	6	0	0	0
6_0_0_0	990.327	11.7	6	0	0	0
6_0_0_0	990.327	12.7	6	0	0	0
6_0_0_0	990.327	14.3	6	0	0	0
6_0_0_0	990.327	15.3	6	0	0	0
6_0_0_0	990.327	25.5	6	0	0	0
6_0_0_0	990.327	27.6	6	0	0	0
6_0_0_0	990.327	28.8	6	0	0	0
6_0_0_0	990.327	31.5	6	0	0	0
6_0_0_0	990.327	33.9	6	0	0	0

4_0_3_0	1062.349	9.7	4	0	3	0
0_0_8_0	1074.349	23.4	0	0	8	0
0_0_8_0	1074.349	24.7	0	0	8	0
0_0_8_0	1074.349	26.6	0	0	8	0
0_0_8_0	1074.349	31.2	0	0	8	0
0_0_8_0	1074.349	32.5	0	0	8	0
0_0_8_0	1074.349	34.3	0	0	8	0
7_0_0_0	1152.380	13.9	7	0	0	0
7_0_0_0	1152.380	15.1	7	0	0	0
0_0_9_0	1206.391	23.5	0	0	9	0
0_0_9_0	1207.399	26.5	0	0	9	0
8_0_0_0	1314.433	15.0	8	0	0	0
0_0_10_0	1320.423	26.3	0	0	10	0
4_0_5_0	1326.433	22.9	4	0	5	0
9_0_0_0	1476.486	16.3	9	0	0	0
10_0_0_0	1638.539	17.3	10	0	0	0
11_0_0_0	1800.591	18.9	11	0	0	0
12_0_0_0	1962.644	19.0	12	0	0	0

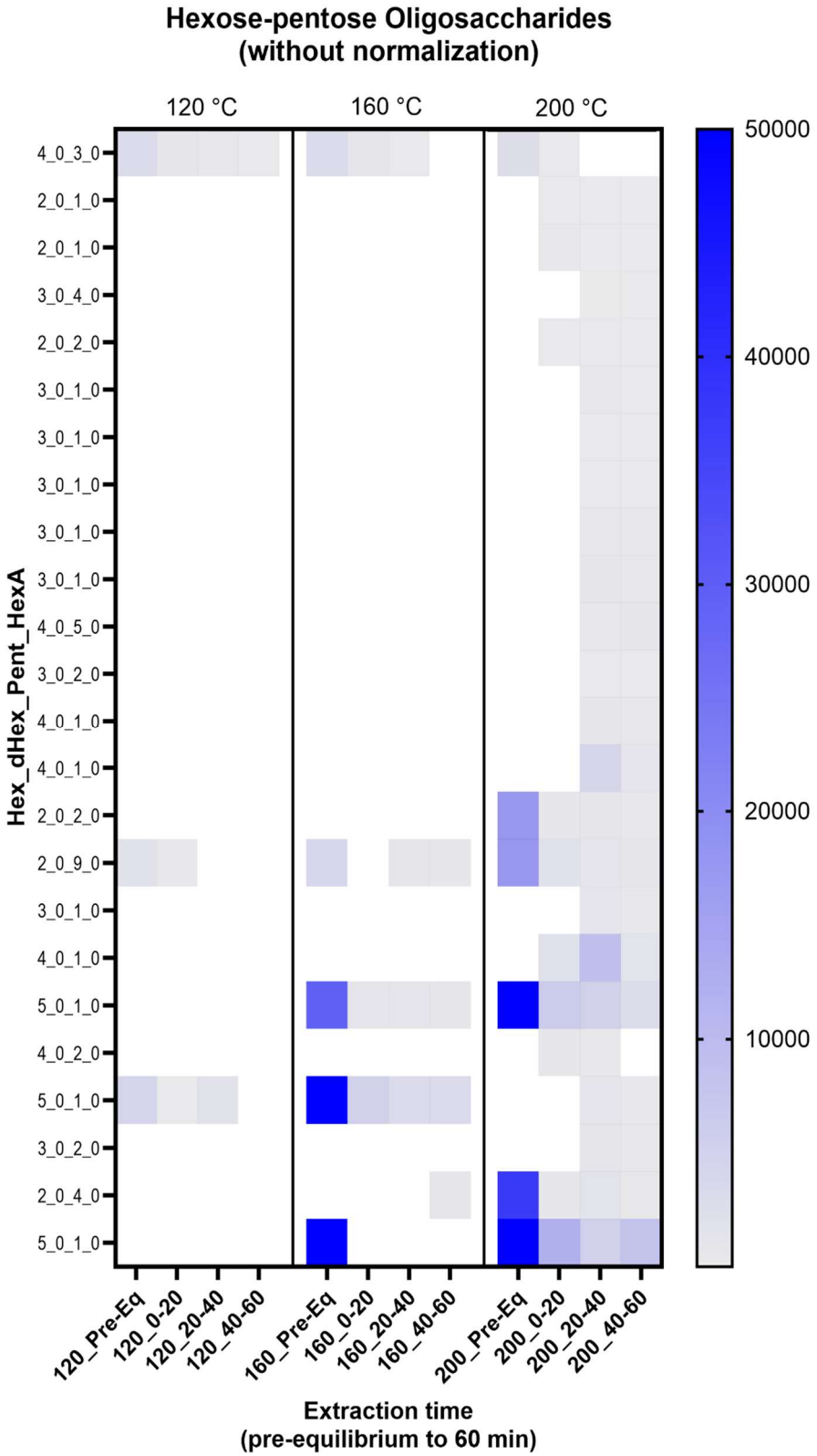
A)


Hexose-only Oligosaccharides
(without normalization)
120 °C
160 °C
200 °C
Hex_dHex_Pent_HexA
3_0_0_0
4_0_0_0
5_0_0_0
6_0_0_0
7_0_0_0
8_0_0_0
9_0_0_0
10_0_0_0
11_0_0_0
12_0_0_0
500000
400000
300000
200000
100000
120_Pre-Eq
120_0-20
120_20-40
120_40-60
160_Pre-Eq
160_0-20
160_20-40
160_40-60
200_Pre-Eq
200_0-20
200_20-40
200_40-60
Extraction time
(pre-equilibrium to 60 min)

B)


Pentose-only Oligosaccharides
(without normalization)
120 °C
160 °C
200 °C
Hex_dHex_Pent_HexA
0_0_3_0
0_0_4_0
0_0_4_0
0_0_4_0
0_0_4_0
0_0_4_0
0_0_4_0
0_0_4_0
0_0_4_0
0_0_4_0
0_0_5_0
0_0_5_0
0_0_5_0
0_0_5_0
0_0_5_0
0_0_5_0
0_0_6_0
0_0_6_0
0_0_6_0
0_0_6_0
0_0_6_0
0_0_7_0
0_0_7_0
0_0_7_0
0_0_7_0
0_0_7_0
0_0_8_0
0_0_8_0
0_0_8_0
0_0_9_0
0_0_9_0
500000
400000
300000
200000
100000
120_Pre-Eq
120_0-20
120_20-40
120_40-60
160_Pre-Eq
160_0-20
160_20-40
160_40-60
200_Pre-Eq
200_0-20
200_20-40
200_40-60
Extraction time
(pre-equilibrium to 60 min)

C)


Hexose-pentose Oligosaccharides
(without normalization)
120 °C
160 °C
200 °C
Hex_dHex_Pent_HexA
4_0_3_0
2_0_1_0
2_0_1_0
3_0_4_0
2_0_2_0
3_0_1_0
3_0_1_0
3_0_1_0
3_0_1_0
3_0_1_0
4_0_5_0
3_0_2_0
4_0_1_0
4_0_1_0
2_0_2_0
2_0_9_0
3_0_1_0
4_0_1_0
5_0_1_0
4_0_2_0
5_0_1_0
3_0_2_0
2_0_4_0
5_0_1_0
50000
40000
30000
20000
10000
120_Pre-Eq
120_0-20
120_20-40
120_40-60
160_Pre-Eq
160_0-20
160_20-40
160_40-60
200_Pre-Eq
200_0-20
200_20-40
200_40-60
Extraction time
(pre-equilibrium to 60 min)

Supplementary Figures S3.1 A-C. Glycoprofile heatmaps of lentil insoluble subcritical water extractions, without soluble carbohydrate content normalization. Oligosaccharides are categorized as containing hexose-only (A), pentose-only (B), and hexose-pentose (C). Individual oligosaccharides are reported on each row of the heatmaps, and the four-digit codes along the left y-axis indicate the number of monosaccharides making up the oligosaccharide (Hex_dHex_Pent_HexA). The right color legend indicates peak areas. Oligosaccharides for each subcritical temperature (120 °C, 160 °C, 200 °C) and extraction time (pre-equilibrium, 0 - 20 min, 20 - 40 min, and 40 - 60 min) are reported.

References

Abdel-Gawad, A. (1993). Effect of domestic processing on oligosaccharide content of some dry legume seeds. *Food Chemistry*, *46*(1), 25–31. https://doi.org/10.1016/0308-8146(93)90070-V

Agrawal, G. K., Sarkar, A., Righetti, P. G., Pedreschi, R., Carpentier, S., Wang, T., … Rakwal, R. (2013). A decade of plant proteomics and mass spectrometry: translation of technical advancements to food security and safety issues. *Mass Spectrometry Reviews*, *32*(5), 335–365. https://doi.org/10.1002/mas.21365

Almeida, J. C. de, Perassolo, M. S., Camargo, J. L., Bragagnolo, N., and Gross, J. L. (2006). Fatty acid composition and cholesterol content of beef and chicken meat in Southern Brazil. *Revista Brasileira de Ciências Farmacêuticas*, *42*(1). https://doi.org/10.1590/S1516-93322006000100012

Alonso-Miravalles, L., Jeske, S., Bez, J., Detzel, A., Busch, M., Krueger, M., … Arendt, E. K. (2019). Membrane filtration and isoelectric precipitation technological approaches for the preparation of novel, functional and sustainable protein isolate from lentils. *European Food Research and Technology = Zeitschrift Fur Lebensmittel-Untersuchung Und -Forschung. A*, *245*(9), 1855–1869. https://doi.org/10.1007/s00217-019-03296-y

Armaforte, E., Hopper, L., and Stevenson, G. (2021). Preliminary investigation on the effect of proteins of different leguminous species (Cicer arietinum, Vicia faba and Lens culinarius) on the texture and sensory properties of egg-free mayonnaise. *LWT*, *136*, 110341. https://doi.org/10.1016/j.lwt.2020.110341

Asakereh, A., Shiekhdavoodi, M. J., and Safaieenejad, M. (2010). *Energy consumption pattern of organic and conventional lentil in Iran a case study: Kuhdasht County*.

Avigad, G., and Dey, P. M. (1997). Storage carbohydrates. *Plant Biochemistry*, 143–204.

Berrios, J. D. J., Morales, P., Cámara, M., and Sánchez-Mata, M. C. (2010). Carbohydrate composition of raw and extruded pulse flours. *Food Research International*, *43*(2), 531–536. https://doi.org/10.1016/j.foodres.2009.09.035

Bhatty, R. S., Slinkard, A. E., and Sosulski, F. W. (1976). Chemical composition and protein characteristics of lentils. *Canadian Journal of Plant Science*, *56*(4), 787–794. https://doi.org/10.4141/cjps76-128

Boeck, T., Zannini, E., Sahin, A. W., Bez, J., and Arendt, E. K. (2021). Nutritional and Rheological Features of Lentil Protein Isolate for Yoghurt-Like Application. *Foods*, *10*(8). https://doi.org/10.3390/foods10081692

Boye, J., Zare, F., and Pletch, A. (2010). Pulse proteins: Processing, characterization, functional properties and applications in food and feed. *Food Research International*, *43*(2), 414–431. https://doi.org/10.1016/j.foodres.2009.09.003

Bravo, L., Siddhuraju, P., and Saura-Calixto, F. (1998). Effect of various processing methods on the in vitro starch digestibility and resistant starch content of indian pulses. *Journal of Agricultural and Food Chemistry*, *46*(11), 4667–4674. https://doi.org/10.1021/jf980251f

Brummer, Y., Kaviani, M., and Tosh, S. M. (2015). Structural and functional characteristics of dietary fibre in beans, lentils, peas and chickpeas. *Food Research International*, *67*, 117–125. https://doi.org/10.1016/j.foodres.2014.11.009

Buckeridge, M. S., Pessoa dos Santos, H., and Tiné, M. A. S. (2000). Mobilisation of storage cell wall polysaccharides in seeds. *Plant Physiology and Biochemistry*, *38*(1–2), 141–156. https://doi.org/10.1016/S0981-9428(00)00162-5

Cai, R., McCurdy, A., and Baik, B. K. (2002). Textural Property of 6 Legume Curds in Relation to their Protein Constituents. *Journal of Food Science*, *67*(5), 1725–1730. https://doi.org/10.1111/j.1365-2621.2002.tb08713.x

Campbell, Kerry A, Vaca-Medina, G., Glatz, C. E., and Pontalier, P.-Y. (2016). Parameters affecting enzyme-assisted aqueous extraction of extruded sunflower meal. *Food Chemistry*, *208*, 245–251. https://doi.org/10.1016/j.foodchem.2016.03.098

Campbell, K A, and Glatz, C. E. (2009). Mechanisms of aqueous extraction of soybean oil. *Journal of Agricultural and Food Chemistry*, *57*(22), 10904–10912. https://doi.org/10.1021/jf902298a

Chang, L., Gu, Z., Bandillo, N., Chen, B., and Rao, J. (2023). Fractionation, Structural Characteristics, Functionality, Aromatic Profile, and In Vitro Digestibility of Lentil (*Lens culinaris*) Proteins. *ACS Food Science & Technology*, *3*(3), 478–488. https://doi.org/10.1021/acsfoodscitech.2c00429

Chilomer, K., Zaleska, K., Ciesiolka, D., Gulewicz, P., Frankiewicz, A., and Gulewicz, K. (2010). Changes in the alkaloid, alpha-galactoside and protein fractions content during germination of different lupin species. *Acta Societatis Botanicorum Poloniae*.

Chiou, T.-Y., Neoh, T. L., Kobayashi, T., and Adachi, S. (2011). Antioxidative ability of defatted rice bran extract obtained by subcritical water extraction in bulk oil and aqueous dispersion systems. *Japan Journal of Food Engineering*, *12*(4), 147–154. https://doi.org/10.11301/jsfe.12.147

de Almeida Costa, G. E., da Silva Queiroz-Monici, K., Pissini Machado Reis, S. M., and de Oliveira, A. C. (2006). Chemical composition, dietary fibre and resistant starch contents of raw and cooked pea, common bean, chickpea and lentil legumes. *Food Chemistry*, *94*(3), 327–330. https://doi.org/10.1016/j.foodchem.2004.11.020

de Moura, J. M. L. N., Campbell, K., Mahfuz, A., Jung, S., Glatz, C. E., and Johnson, L. (2008). Enzyme-Assisted Aqueous Extraction of Oil and Protein from Soybeans and Cream De-emulsification. *Journal of the American Oil Chemists' Society*, *85*(10), 985–995. https://doi.org/10.1007/s11746-008-1282-2

Dhingra, D., Michael, M., Rajput, H., and Patil, R. T. (2012). Dietary fibre in foods: a review. *Journal of Food Science and Technology*, *49*(3), 255–266. https://doi.org/10.1007/s13197-011-0365-5

Dhull, S. B., Kinabo, J., and Uebersax, M. A. (2023). Nutrient profile and effect of processing methods on the composition and functional properties of lentils (Medik): A review. *Legume*

Science, *5*(1). https://doi.org/10.1002/leg3.156

Dias, F. F. G., Yang, J. S., Pham, T. T. K., Barile, D., and de Moura Bell, J. M. L. N. (2023). Unveiling the contribution of Osborne protein classes to alkaline and enzymatically extracted green lentil proteins. Manuscript in preparation.

DuBois, Michel., Gilles, K. A., Hamilton, J. K., Rebers, P. A., and Smith, Fred. (1956). Colorimetric Method for Determination of Sugars and Related Substances. *Analytical Chemistry*, *28*(3), 350–356. https://doi.org/10.1021/ac60111a017

Eckert, E., Wismer, W., Waduthanthri, K., Babii, O., Yang, J., and Chen, L. (2018). Application of Barley- and Lentil-Protein Concentrates in the Production of Protein-Enriched Doughnuts. *Journal of the American Oil Chemists' Society*, *95*(8), 1027–1040. https://doi.org/10.1002/aocs.12103

El-Nahry, F. I., Mourad, F. E., Abdel Khalik, S. M., and Bassily, N. S. (1980). Chemical composition and protein quality of lentils (Lens) consumed in Egypt. *Plant Foods for Human Nutrition*, *30*(2), 87–95. https://doi.org/10.1007/BF01099046

FAO Food and Agriculture Organization of the United Nations. (2023). FAOSTAT Statistical Database. Retrieved July 10, 2023, from Food and Agriculture Statistical Database website: https://www.fao.org/food-agriculture-statistics/en/

Fekadu Gemede, H. (2014). Antinutritional factors in plant foods: potential health benefits and adverse effects. *International Journal of Nutrition and Food Sciences*, *3*(4), 284. https://doi.org/10.11648/j.ijnfs.20140304.18

Finley, J. W., Burrell, J. B., and Reeves, P. G. (2007). Pinto bean consumption changes SCFA profiles in fecal fermentations, bacterial populations of the lower bowel, and lipid profiles in blood of humans. *The Journal of Nutrition*, *137*(11), 2391–2398. https://doi.org/10.1093/jn/137.11.2391

Fleming, S. E. (1981). A study of relationships between flatus potential and carbohydrate distribution in legume seeds. *Journal of Food Science*, *46*(3), 794–798. https://doi.org/10.1111/j.1365-2621.1981.tb15350.x

Ghumman, A., Kaur, A., and Singh, N. (2016). Functionality and digestibility of albumins and globulins from lentil and horse gram and their effect on starch rheology. *Food Hydrocolloids*, *61*, 843–850. https://doi.org/10.1016/j.foodhyd.2016.07.013

Gilmour, S. J., Sebolt, A. M., Salazar, M. P., Everard, J. D., and Thomashow, M. F. (2000). Overexpression of the Arabidopsis CBF3 transcriptional activator mimics multiple biochemical changes associated with cold acclimation. *Plant Physiology*, *124*(4), 1854–1865. https://doi.org/10.1104/pp.124.4.1854

Hall, C., Hillen, C., and Garden Robinson, J. (2017). Composition, nutritional value, and health benefits of pulses. *Cereal Chemistry Journal*, *94*(1), 11–31. https://doi.org/10.1094/CCHEM-03-16-0069-FI

Hanmoungjai, P., Pyle, D. L., and Niranjan, K. (2002). Enzyme-assisted water-extraction of oil and protein from rice bran. *Journal of Chemical Technology & Biotechnology*, *77*(7), 771–

776. https://doi.org/10.1002/jctb.635

Huang, Y.-P., Robinson, R. C., and Barile, D. (2022). Food glycomics: Dealing with unexpected degradation of oligosaccharides during sample preparation and analysis. *Journal of Food and Drug Analysis*, *30*(1), 62–76. https://doi.org/10.38212/2224-6614.3393

Ialicicco, M., Viscosi, V., Arena, S., Scaloni, A., Trupiano, D., Rocco, M., … Scippa, G. S. (2012). Lens culinaris Medik. seed proteome: analysis to identify landrace markers. *Plant Science*, *197*, 1–9. https://doi.org/10.1016/j.plantsci.2012.08.010

Jarpa-Parra, M., Wong, L., Wismer, W., Temelli, F., Han, J., Huang, W., Chen, L. (2017). Quality characteristics of angel food cake and muffin using lentil protein as egg/milk replacer. *International Journal of Food Science & Technology*, *52*(7), 1604–1613. https://doi.org/10.1111/ijfs.13433

Jarpa-Parra, M. (2018). Lentil protein: a review of functional properties and food application. An overview of lentil protein functionality. *International Journal of Food Science & Technology*, *53*(4), 892–903. https://doi.org/10.1111/ijfs.13685

Joshi, M., Timilsena, Y., and Adhikari, B. (2017). Global production, processing and utilization of lentil: A review. *Journal of Integrative Agriculture*, *16*(12), 2898–2913. https://doi.org/10.1016/S2095-3119(17)61793-3

Karaca, A. C., Low, N., and Nickerson, M. (2011). Emulsifying properties of chickpea, faba bean, lentil and pea proteins produced by isoelectric precipitation and salt extraction. *Food Research International*, *44*(9), 2742–2750. https://doi.org/10.1016/j.foodres.2011.06.012

Kataoka, M., Wiboonsirikul, J., Kimura, Y., and Adachi, S. (2008). Properties of Extracts from Wheat Bran by Subcritical Water Treatment. *Food Science and Technology Research*, *14*(6), 553–556. https://doi.org/10.3136/fstr.14.553

Kimura, A., Fukuda, T., Zhang, M., Motoyama, S., Maruyama, N., and Utsumi, S. (2008). Comparison of physicochemical properties of 7S and 11S globulins from pea, fava bean, cowpea, and French bean with those of soybean--French bean 7S globulin exhibits excellent properties. *Journal of Agricultural and Food Chemistry*, *56*(21), 10273–10279. https://doi.org/10.1021/jf801721b

Luna-Vital, D. A., Mojica, L., González de Mejía, E., Mendoza, S., and Loarca-Piña, G. (2015). Biological potential of protein hydrolysates and peptides from common bean (Phaseolus vulgaris L.): A review. *Food Research International*, *76*, 39–50. https://doi.org/10.1016/j.foodres.2014.11.024

Marlett, J. A., McBurney, M. I., and Slavin, J. L. (2002). Position of the american dietetic association. *Journal of the American Dietetic Association*, *102*(7), 993–1000. https://doi.org/10.1016/S0002-8223(02)90228-2

Matny, O. N. (2015). Lentil (lens culinaris medikus) current status and future prospect of production in ethiopia. *Advances in Plants & Agriculture Research*, *2*(2). https://doi.org/10.15406/apar.2015.02.00040

Ma, Z., Boye, J. I., and Simpson, B. K. (2016). Preparation of salad dressing emulsions using lentil, chickpea and pea protein isolates: A response surface methodology study. *Journal of*

Food Quality, *39*(4), 274–291. https://doi.org/10.1111/jfq.12190

Machida, K., Huang, Y.-P., Furlan Gonçalves Dias, F., Barile, D., and de Moura Bell, J. M.L. N. (2022). Leveraging Bioprocessing Strategies to Achieve the Simultaneous Extraction of Full-Fat Chickpea Flour Macronutrients and Enhance Protein and Carbohydrate Functionality. *Food and Bioprocess Technology*, *15*, 1760–1777. https://doi.org/10.1007/s11947-022-02847-8

National Center for Biotechnology Information. (n.d.). Protein - NCBI. Retrieved July 17, 2023, from NCBI Protein Database website: https://www.ncbi.nlm.nih.gov/protein/?term=

Neves, V. A., and Lourenço, E. J. (1995). Isolation and in vitro hydrolysis of globulin g1 from lentils (lens culinaris, medik). *Journal of Food Biochemistry*, *19*(2), 109–120. https://doi.org/10.1111/j.1745-4514.1995.tb00525.x

Obendorf, R. L. (1997). Oligosaccharides and galactosyl cyclitols in seed desiccation tolerance. *Seed Science Research*, *7*(2), 63–74. https://doi.org/10.1017/S096025850000341X

Ogun, P. O., Markakis, P., and Chenoweth, W. (1989). Effect of Processing on Certain Antinutrients in Cowpeas (Vigna unguiculata). *Journal of Food Science*, *54*(4), 1084–1085. https://doi.org/10.1111/j.1365-2621.1989.tb07952.x

Osborne, T. B. (1924). The Vegetable Proteins. *Nature*, *114*(2875), 822–822. https://doi.org/10.1038/114822c0

Osemwota, E. C., Alashi, A. M., and Aluko, R. E. (2022). Physicochemical and functional properties of albumin, globulin and glutelin fractions of green lentil seed. *International Journal of Food Science & Technology*, *57*(7), 3967–3981. https://doi.org/10.1111/ijfs.15608

Perera, A., Meda, V., and Tyler, R. T. (2010). Resistant starch: A review of analytical protocols for determining resistant starch and of factors affecting the resistant starch content of foods. *Food Research International*, *43*(8), 1959–1974. https://doi.org/10.1016/j.foodres.2010.06.003

Price, K. R., Lewis, J., Wyatt, G. M., and Fenwick, G. R. (1988). Flatulence--causes, relation to diet and remedies. *Die Nahrung*, *32*(6), 609–626. https://doi.org/10.1002/food.19880320626

Ramirez, C. S. V., Temelli, F., and Saldaña, M. D. A. (2021). Production of pea hull soluble fiber-derived oligosaccharides using subcritical water with carboxylic acids. *The Journal of Supercritical Fluids*, 105349. https://doi.org/10.1016/j.supflu.2021.105349

Reddy, A. A. (2004, October 30). Consumption Pattern, Trade and Production Potential of Pulses. *Economic and Political Weekly*, *39*, 4854–4860.

Rizkalla, S. W., Bellisle, F., and Slama, G. (2002). Health benefits of low glycaemic index foods, such as pulses, in diabetic patients and healthy individuals. *The British Journal of Nutrition*, *88 Suppl 3*, S255-62. https://doi.org/10.1079/BJN2002715

Romano, A., Gallo, V., Ferranti, P., and Masi, P. (2021). Lentil flour: nutritional and

technological properties, in vitro digestibility and perspectives for use in the food industry. *Current Opinion in Food Science*, *40*, 157–167. https://doi.org/10.1016/j.cofs.2021.04.003

Samaranayaka, A. (2017). Lentil. In *Sustainable Protein Sources* (pp. 185–196). Elsevier. https://doi.org/10.1016/B978-0-12-802778-3.00011-1

Sathe, S. K. (2002). Dry bean protein functionality. *Critical Reviews in Biotechnology*, *22*(2), 175–223. https://doi.org/10.1080/07388550290789487

Scippa, G. S., Rocco, M., Ialicicco, M., Trupiano, D., Viscosi, V., Di Michele, M., … Scaloni, A. (2010). The proteome of lentil (Lens culinaris Medik.) seeds: discriminating between landraces. *Electrophoresis*, *31*(3), 497–506. https://doi.org/10.1002/elps.200900459

Shevkani, K., Singh, N., Chen, Y., Kaur, A., and Yu, L. (2019). Pulse proteins: secondary structure, functionality and applications. *Journal of Food Science and Technology*, *56*(6), 2787–2798. https://doi.org/10.1007/s13197-019-03723-8

Shewry, P. R., Napier, J. A., and Tatham, A. S. (1995). Seed storage proteins: structures and biosynthesis. *The Plant Cell*, *7*(7), 945–956. https://doi.org/10.1105/tpc.7.7.945

Shewry, P. R. (1995). Plant storage proteins. *Biological Reviews of the Cambridge Philosophical Society*, *70*(3), 375–426. https://doi.org/10.1111/j.1469-185x.1995.tb01195.x

Shrestha, S., van ’t Hag, L., Haritos, V., and Dhital, S. (2023). Comparative study on molecular and higher-order structures of legume seed protein isolates: Lentil, mungbean and yellow pea. *Food Chemistry*, *411*, 135464. https://doi.org/10.1016/j.foodchem.2023.135464

Sidhu, J. S., Zafar, T., Benyathiar, P., and Nasir, M. (2022). Production, processing, and nutritional profile of chickpeas and lentils. In M. Siddiq and M. A. Uebersax (Eds.), *Dry beans and pulses: production, processing, and nutrition* (pp. 383–407). Wiley. https://doi.org/10.1002/9781119776802.ch15

Silva-Cristobal, L., Osorio-Díaz, P., Tovar, J., and Bello-Pérez, L. A. (2010). Chemical composition, carbohydrate digestibility, and antioxidant capacity of cooked black bean, chickpea, and lentil Mexican varieties Composición química, digestibilidad de carbohidratos, y capacidad antioxidante de variedades mexicanas cocidas de frijol negro, garbanzo, y lenteja. *CyTA - Journal of Food*, *8*(1), 7–14. https://doi.org/10.1080/19476330903119218

Siva, N., Thavarajah, D., Johnson, C. R., Duckett, S., Jesch, E. D., and Thavarajah, P. (2017). Can lentil (Lens culinaris Medikus) reduce the risk of obesity? *Journal of Functional Foods*. https://doi.org/10.1016/j.jff.2017.02.017

Siva, N., Thavarajah, P., and Thavarajah, D. (2018). The impact of processing and cooking on prebiotic carbohydrates in lentil. *Journal of Food Composition and Analysis*, *70*, 72–77. https://doi.org/10.1016/j.jfca.2018.04.006

Souza Almeida, F., Furlan Goncalves Dias, F., Kawazoe Sato, A. C., and Leite Nobrega de Moura Bell, J. M. (2021). From solvent extraction to the concurrent extraction of lipids and proteins from green coffee: An eco-friendly approach to improve process feasibility. *Food and Bioproducts Processing*, *129*, 144–156. https://doi.org/10.1016/j.fbp.2021.08.004

Souza, T. S. P., Dias, F. F. G., Koblitz, M. G. B., and M. L. N. de M. Bell, J. (2019). Aqueous and Enzymatic Extraction of Oil and Protein from Almond Cake: A Comparative Study. *Processes*, *7*(7), 472. https://doi.org/10.3390/pr7070472

Tahir, M., Lindeboom, N., Båga, M., Vandenberg, A., and Chibbar, R. (2011). Composition and correlation between major seed constituents in selected lentil (*Lens culinaris.* Medik) genotypes. *Canadian Journal of Plant Science*, *91*(5), 825–835. https://doi.org/10.4141/cjps2011-010

Tovar, J., and Melito, C. (1996). Steam-Cooking and Dry Heating Produce Resistant Starch in Legumes. *Journal of Agricultural and Food Chemistry*, *44*(9), 2642–2645. https://doi.org/10.1021/jf950824d

U.S. Department of Agriculture, Agricultural Research Service. (2019). FoodData Central. Retrieved July 14, 2023, from FoodData Central website: https://fdc.nal.usda.gov/

Verde, C. V.-V., Frías, J., and Verde, S. V. (1992). Effect of processing on the soluble carbohydrate content of lentils. *Journal of Food Protection*, *55*(4), 301–303. https://doi.org/10.4315/0362-028X-55.4.301

Vidal-Valverde, C., Frías, J., and Valverde, S. (1993). Changes in the carbohydrate composition of legumes after soaking and cooking. *Journal of the American Dietetic Association*, *93*(5), 547–550. https://doi.org/10.1016/0002-8223(93)91814-7

Wang, N., and Daun, J. (2006). Effects of variety and crude protein content on nutrients and anti-nutrients in lentils (). *Food Chemistry*, *95*(3), 493–502. https://doi.org/10.1016/j.foodchem.2005.02.001

Wang, X., Cai, X., Xu, C., Wang, Q., and Dai, S. (2016). Drought-Responsive Mechanisms in Plant Leaves Revealed by Proteomics. *International Journal of Molecular Sciences*, *17*(10). https://doi.org/10.3390/ijms17101706

Wiboonsirikul, J., Hata, S., Tsuno, T., Kimura, Y., and Adachi, S. (2007). Production of functional substances from black rice bran by its treatment in subcritical water. *LWT - Food Science and Technology*, *40*(10), 1732–1740. https://doi.org/10.1016/j.lwt.2007.01.003

Wiboonsirikul, J., Mori, M., Khuwijitjaru, P., and Adachi, S. (2013). Properties of Extract from Okara by Its Subcritical Water Treatment. *International Journal of Food Properties*, *16*(5), 974–982. https://doi.org/10.1080/10942912.2011.573119

Wright, J. C., Beynon, R. J., and Hubbard, S. J. (2010). Cross species proteomics. *Methods in Molecular Biology*, *604*, 123–135. https://doi.org/10.1007/978-1-60761-444-9_9

Yang, J. S., Dias, F. F. G., Pham, T. T. K., Barile, D., and de Moura Bell, J. M. L. N. (2023). A sequential fractionation approach to understanding the physicochemical and functional properties of aqueous and enzyme-assisted aqueous extracted black bean proteins. *Food Hydrocolloids*. Submitted for publication.

CHAPTER 4:

Algae - giant kelp (*Macrocystis pyrifera*) protein and carbohydrate characterization and elucidating the impact of green processes

A modified version of this chapter was submitted as a report in the Department of Energy AlgaePrize competition.[3]

Abstract

The utilization of macroalgae, specifically the giant kelp *Macrocystis pyrifera*, as a rich source of functional and bioactive compounds has drawn significant interest across various industrial applications. However, the lack of effective and environmentally friendly green extraction methods for obtaining these valuable compounds from giant kelp remains a critical research gap. This study aims to bridge this gap by using proteomics and glycomics techniques to evaluate the efficacy of different green extraction methods, including aqueous, enzyme-assisted, microwave-assisted, and microwave-enzyme approaches, by analyzing the protein and carbohydrate compositions resulting from these extraction technologies. The use of analytical techniques can explain the composition of the extracts and therefore their biological activities. Proteomics analysis enabled the identification of proteins released during aqueous extraction, shedding light on the impact of different extraction conditions on protein composition. Surprisingly, only nine proteins were identified in aqueous extracts, and proteomics data exhibited discrepancies when compared to other protein analysis results, highlighting potential challenges in sample preparation or analysis. Nevertheless, these insights serve as valuable stepping stones for future investigations into unlocking the full potential of giant kelp proteins. Carbohydrate analysis, using glycomics and monosaccharide analysis, provided crucial information on the oligosaccharides and polysaccharides released by various green extraction methods. Glycomics analysis investigated the oligosaccharides released through these various extraction methods, revealing minor differences between 6 h of aqueous extraction and 15 min of microwave-assisted extraction. Additionally, microwave-assisted extraction demonstrated

[3] Gaiero, M., Pham, T. T., Barile, D., and de Moura Bell, J. M. L. N. (2023). Kelping it Green: Innovative Analytical Tools for Eco-Friendly Giant Kelp Processing. U.S. Department of Energy AlgaePrize Competition.

less dependence on pH, making it a more robust and promising green extraction technology for industrial applications, especially considering other compounds, such as proteins, could be optimized alongside carbohydrates. Similarly, monosaccharide analysis, employed to quantify the monosaccharide constituents of fucoidan, a polysaccharide of interest, found comparable monosaccharide contents between 6 h aqueous extraction and 15 min microwave-assisted extraction at neutral and basic pH levels. These results signify the potential of microwave-assisted extraction as a viable and efficient green extraction technique for industrial applications. Challenges arose due to unexpected indigenous carbohydrates which were undeclared on the enzyme label, impacting the acquisition of accurate quantitative carbohydrate data. However, this glycomics approach remains instrumental in understanding the effects of green extraction technologies on the carbohydrate profile of giant kelp. Overall, this integrative approach of proteomics and glycomics analysis provided insights into the impact of green extraction methods on the proteins and carbohydrates present in giant kelp and provided hints towards elucidating their functions. This study contributes to the advancement of effective and sustainable techniques for producing functional and bioactive compounds from giant kelp. These findings hold promise for applications in pharmaceuticals, nutraceuticals, and other industries, ultimately unlocking the valuable potential of giant kelp.

Introduction

Algae, a group of plant-like organisms, are abundant in various aquatic and terrestrial environments (PhycoTerra, 2022). The term “algae” encompasses a diverse group of species, but algae can broadly be categorized as microalgae, which are single-celled, and macroalgae, which are multicellular and commonly found in the ocean (PhycoTerra, 2022). Macroalgae are further classified into red (Rhodophyta), green (Chlorophyta), and brown (Ochrophyta) algae (Pereira, 2021). Brown algae are particularly notable for their rapid growth and diverse range of compounds with valuable nutritional, biological, and functional properties (Garcia-Vaquero et al.,

2017; Yong et al., 2022). In 2019, brown algae alone accounted for over 17 million tons of global algae production (FAO of the UN, 2021).

Macroalgae have been consumed in Asian diets for centuries, however more recently, increasing interest in sustainable foods and macroalgae-derived products have gained more attention (Dobrinčić et al., 2020; Yong et al., 2022). The United Nations Framework Convention on Climate Change (UNFCCC) has recognized macroalgae as an important tool for achieving carbon-cutting goals, promoting carbon sequestration and organic carbon storage in marine ecosystems (Yong et al., 2022). Compared to terrestrial crops, macroalgae can grow rapidly in aquatic environments (Dobrinčić et al., 2020), do not compete for land or fertilization resources, and have the potential to absorb approximately 20 times more carbon dioxide per acre than land-based forests, making their cultivation a powerful strategy for mitigating climate change and providing environmental and societal benefits, particularly for coastal communities (Global Market Insights, 2021; IMARC group, 2021).

Macroalgae can be cultivated through both wild and farmed methods. While wild production has remained relatively stable, macroalgae farming has experienced significant growth in recent years. The Food and Agricultural Organization and the United Nations have reported a remarkable increase of over 6000% in global macroalgae production from 1950 to 2019, with approximately 36 million tonnes produced in 2019, and this trend is projected to continue (FAO of the UN, 2021). Despite oceans covering 71% of the Earth, only 2% of our global food supply is sourced from the sea, primarily through seaweed aquaculture concentrated in Asian countries, particularly China (Lloyds Register Foundation). To ensure sustainable growth and meet the demands of a growing population, it is imperative to expand marine aquaculture globally and avoid reaching the biophysical limits of available growing spaces (Duarte et al., 2021).

Seaweed compounds have found applications in various fields, including pharmaceutical drug delivery systems, bioplastic production, thickening agents, dietary supplements, and animal feed (Duarte et al., 2021; Yong et al., 2022). Nevertheless, there is still much to learn about the functional and biological properties of these compounds. Ongoing research efforts aim to discover new uses and applications for the diverse compounds found in macroalgae in order to maximize the benefits derived from macroalgae production.

A major challenge in the development of green extraction methods lies in comprehensively understanding the effects of different extraction techniques and key processing conditions on the extractability and structure-function relationship of the algae compounds. It is essential to employ analytical methods capable of elucidating the impact of these conditions on the complex structure of bioactive compounds present in macroalgae. By doing so, structure-function based methods can be developed to optimize the extraction process and generate compounds with the desired functionality. The use of analytical techniques can explain the composition of the extracts and therefore their biological activities.

Algae composition

Macrocystis Pyrifera, or giant kelp, is a brown macroalgae species commonly found in the Pacific Ocean containing a diverse array of compounds with desirable properties. The exact composition of giant kelp vary depending on factors such as seasonal variations and geographical location, but they are generally composed of 34-76% carbohydrates, 1-27% proteins, 0.5 - 3.5% lipids, and 9 - 41% ash on a dry basis (Kostas et al., 2021). For the giant kelp used in this study, the proximate composition was determined to be 5.06 ± 0.45% moisture, 1.80 ± 0.10% lipids, 23.16 ± 0.11% ash, 6.51 ± 0.05% protein, and 63.4% carbohydrate (Gaiero et al., 2023).

Algae proteins

The protein content in giant kelp is relatively lower (1-27%) compared to carbohydrates (Kostas et al., 2021). The majority of proteins in macroalgae are produced intracellularly, and their extractability depends on the selection of extraction conditions that promote protein solubility and diffusion from the algal matrix into the extraction medium (Gordalina et al., 2021). The literature concerning protein extraction and characterization in brown macroalgae is limited compared to other crops due to the predominant focus on extracting the more abundant carbohydrates from these organisms (Sari et al., 2015)) and the inherent challenges associated with extracting proteins from macroalgae (Barbarino and Lourenço, 2005). Further, making comparisons between studies is difficult because protein composition is highly dependent on the method of protein extraction used, and algae composition can vary greatly depending on season and geography, even for the same species (Barbarino and Lourenço, 2005; Kostas et al., 2021).

The primary challenge in protein extraction from macroalgae stems from the complex algal matrix. Proteins are embedded in a complex mesh of modified and branched polysaccharides, and the extraction of proteins can be impeded by highly viscous polysaccharides such as alginates (Fleurence, 1999; Wijesinghe and Jeon, 2012). Conventional methods for macroalgae protein extraction include aqueous, acidic, and alkaline methods, and some may not be environmentally sustainable (Bleakley and Hayes, 2017). Therefore, there is a pressing need for the development of green extraction methods that can enhance protein yield.

Despite these challenges, investigating protein extraction and composition in macroalgae remains valuable. Macroalgae contains all the essential amino acids, making it of interest for human consumption (Filote et al., 2021). In fact, certain species of brown macroalgae have

been reported to contain amino acid levels comparable to those found in beef (Marinho et al., 2015). Proteins derived from giant kelp have demonstrated bioactivity, including antioxidant properties (Vásquez et al., 2019). Further research is warranted, as the amino acid content of marine macroalgae holds potential for its utilization as a viable protein source for human consumption and the production of nutraceuticals.

Algae carbohydrates

The carbohydrate fraction of giant kelp includes cellulose and several other polysaccharides of interest such as alginate (10-40% dry basis), fucoidan (5-10% dry basis), and laminarin (22-49% dry basis) (Dobrinčić et al., 2020; Kostas et al., 2021; Michel et al., 2010). The abundance of these polysaccharides varies depending on the species and seasonal factors (Lin et al., 2018). These polysaccharides possess a broad range of functional properties, such as gelling and emulsification, as well as diverse biological activities, including anti-inflammatory, immunostimulatory, antioxidant, antiviral, and antitumor effects (Garcia-Vaquero et al., 2017).

Among these polysaccharides, alginates play a significant role as a major cell wall constituent and primary storage polysaccharide in giant kelp (Filote et al., 2021; Pérez et al., 2016). Alginates are linear polysaccharides composed of mannuronic acid and guluronic acid units (Abraham et al., 2019; Andriamanantoanina and Rinaudo, 2010). They contribute to the flexibility of macroalgae, enabling them to withstand marine currents (Filote et al., 2021). Alginates are currently widely utilized in the food and biomedical industries due to their gelling and emulsification properties (Bordoloi and Goosen, 2020; Filote et al., 2021; Flórez-Fernández et al., 2019).

Fucoidans are a complex and heterogeneous group of sulfated polysaccharides found in the cell walls of brown macroalgae. They consist primarily of linked fucose residues, along with smaller

amounts of other monosaccharides units such as arabinose, glucose, xylose, uronic acid, mannose, and galactose (Bordoloi and Goosen, 2020; Filote et al., 2021). Fucoidan structures are highly variable, encompassing diverse branching patterns, substituents, linkage types, and degrees of sulfation (Pérez et al., 2016). The sulfation of fucoidan is associated with its bioactivity, and studies demonstrate that greater degrees of sulfation correspond to higher biological activity (Filote et al., 2021). Additional factors influencing the biological activity of fucoidan include monosaccharide composition, polymer chain structure, and molecular weight (Atashrazm et al., 2015; Hahn et al., 2012; Saravan et al., 2018; Vishchuk et al., 2011). Notably, fucoidans have demonstrated various beneficial properties, such as antiviral, anti-inflammatory, immunomodulatory, antidiabetic, and anticoagulant effect, among others (Bordoloi and Goosen, 2020; Garcia-Vaquero et al., 2017).

Laminarin are storage polysaccharides predominantly found within cell vacuoles. They are linear beta-glucans (consisting of glucose residues) (Bordoloi and Goosen, 2020; Garcia-Vaquero et al., 2017). Laminarin has exhibited bioactivity, displaying antibacterial, antioxidative, and anticoagulant properties, among others (Bordoloi and Goosen, 2020).

Protein and carbohydrate analysis in the presented work with algae

The primary objective of this work is to develop and utilize protein and carbohydrate analytical methods to guide the development of scalable green downstream processing strategies for macroalgae, specifically the giant kelp species Macrocystis pyrifera. Aqueous, enzyme-assisted, microwave-assisted, and microwave-enzyme extraction methods are utilized with the aim of isolating bioactive compounds from the kelp that possess specific functional and biological properties suitable for potential industrial applications. Analytical techniques to quantify and characterize key biomass compounds (protein, oligosaccharides, monosaccharides) are used to evaluate the impact of sustainable extraction methods on the extractability, structural

composition, and functional/biological properties of giant kelp extracts. This research will facilitate the development of effective and sustainable extraction methods, enabling the production of algae extracts with desired biological and functional properties.

Experimental procedure

Figure 4.1 provides an overview of the experimental procedures involved in the analysis of giant kelp, including extraction and analysis methods.

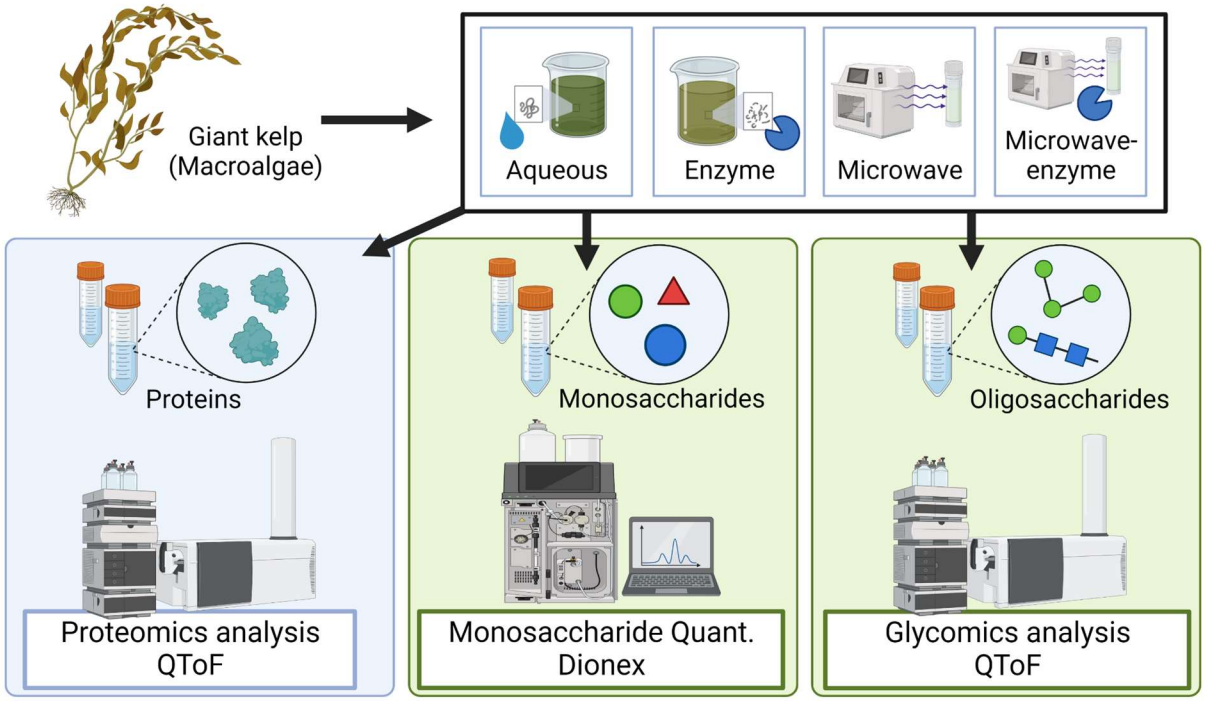


Figure 4.1. Overview of giant kelp project, including extraction (aqueous, enzyme-assisted, microwave-assisted, and microwave-enzyme) and analysis (protein, monosaccharides, and oligosaccharides. Created with BioRender.com

Extraction methods

Aqueous extraction (AEP), enzyme-assisted aqueous extraction (EAEP), microwave-assisted aqueous extraction, and microwave-enzyme aqueous extraction were all investigated. Enzymes used in these processes include carbohydrases, which enables the breakdown of cell wall

cellulose and hemicellulose, thus allowing for greater release of the intracellular compounds, and proteases, which hydrolyze proteins and have been shown to increase extraction yields. Both carbohydrases and proteases have been used in macroalgae extraction (Alboofetileh et al., 2019; Alboofetileh et al., 2018), and proteases have been used to increase the extractability of bioactive proteins (de Moura et al., 2008; Souza Almeida et al., 2021; Vásquez et al., 2019).

Optimal extraction conditions for pH, time, temperature, and solids-to liquids ratio was determined through prior work (Gaiero et al., 2023). FoodPro's acidic carbohydrase (CBL) (5% w/w) and Bio-Cat's Neutral Protease L (NP) (2.5% w/w) were utilized to assist the overall extractability of giant kelp, and the slurry pH was adjusted to values according to each enzyme requirement (pH 4 for CBL and pH 7 for NP). EAEP experiments were performed at 60 °C, under constant stirring of 1400 rpm, for 6 h. All extractions were conducted in triplicate. Following each extraction, extracts were frozen and stored at -10 °C until further analysis.

Proteomics of algae extracts

From the giant kelp extracts, protein was precipitated by adding ice-cold acetone (Fisher Scientific, Waltham, MA)/trichloroacetic acid (MilliporeSigma, St. Louis, MO) (80/20 v/v) containing 5 mM dithiothreitol (Fisher Scientific, Waltham, MA) in a 2:1 ratio (acetone solution:extract). The solution was kept at -20 °C for 1 hour, centrifuged (15,000 x g, 4 °C, 15 min), and the supernatant was removed. The remaining pellet was washed twice with acetone containing 5 mM dithiothreitol, then centrifuged in between washes as before. The pellet was washed a final time with acetone/water (80/20 v/v) containing 5 mM dithiothreitol, centrifuged, and supernatant removed. The pellet was briefly dried under ambient conditions, then weighed to serve as a surrogate measure of total protein mass.

Protein pellets were resuspended in 8M urea (ThermoFisher Scientific, Waltham, MA), 50 mM Tris-HCl (Megazyme, Bray, Ireland), and 5 mM dithiothreitol, vortexed thoroughly, then sonicated in an ice bath (15 30 sec cycles, high intensity). An amount equivalent to 100 µg protein (determined by Qubit 3.0 fluorometer assay, ThermoFisher Scientific, Waltham, MA) was transferred to a fresh tube for digestion. Additional urea solution was added to give a final concentration of 2 µg/µL, and samples were incubated (37 °C, 1 hour). Iodoacetamide (MilliporeSigma, St. Louis, MO) was added to a final concentration of 15 mM and left to alkylate (room temperature, dark, 30 min). Additional dithiothreitol was added to a final concentration of 5 mM to quench excess iodoacetamide. Samples were diluted with 50 mM ammonium bicarbonate (Spectrum Chemical, Gardena, CA) to reduce the concentration of urea to 1 M. Trypsin (Promega, Madison, WI) was added at a 1:40 ratio and incubated (37 °C, overnight) (Eppendorf ThermoMixer C, Enfield, CT). Following overnight incubation, digestion was stopped by adding trifluoroacetic acid (TFA) (MilliporeSigma, St. Louis, MO) until pH 2-3 was reached. Samples were centrifuged (15,000 x g, 4 °C, 15 min) (Eppendorf Centrifuge 5424, Enfield, CT) to remove insoluble particulates. Supernatants containing digested peptides were cleaned by microplate C18 solid phase extraction (Glygen, Columbia, MD). Eluted samples were dried (Genevac miVac, Ipswich, United Kingdom), then reconstituted in 3/97 acetonitrile/water containing 0.1% formic acid.

LC-MS/MS analysis was performed with an Agilent 1200 Series HPLC coupled to an Agilent 6520 Accurate-Mass Q-TOF LC-MS with a Chip Cube interface (Agilent Technologies, Santa Clara, CA) as described by Huang et al (Huang et al., 2022). PEAKS Studio X+ (Bioinformatics Solutions Inc., Waterloo, ON, Canada) was used for peptide and protein identification. Peptides were identified through database search using the UniProt database (https://www.uniprot.org/, accessed 12/08/2022) with the organism name *Macrocystis pyrifera*. The mass error tolerance was 20 ppm and 0.035 Da for the precursor and fragment ions, respectively. The enzyme was

set to “Trypsin” with a specific digestion mode. The number of maximum missed cleavages per peptide was set to 2. Carbamidomethylation was set as a fixed PTM. A maximum of 5 variable modifications, including oxidation, phosphorylation, and deamidation were allowed. The results were filtered with a false discovery rate of 1.0%. Only proteins with at least 1 unique peptide were retained. Database matches were manually inspected to select the correct protein match.

Oligosaccharide profiling of algae extracts

To isolate oligosaccharides, two volumes of ice-cold ethanol were added to extracts, then kept at -20 °C for 1 hour. Samples were centrifuged (4200 x g, 4 °C, 30 min) (Allegra X-30R Centrifuge, Beckman Coulter, Brea, CA). Supernatants were transferred to fresh tubes and dried overnight (Genevac miVac, Ipswich, United Kingdom). The following day, samples were reconstituted in water, with water volumes normalized to carbohydrate content as determined by phenol-sulfuric acid method (DuBois et al., 1956) for all samples. Samples were vortexed, then sonicated (10 min, high intensity) to ensure full dissolution. Samples were centrifuged (14,000 x g, 4 °C, 30 min) (Eppendorf Centrifuge 5424, Enfield, CT) to remove particulates, then cleaned by microplate C18 SPE (Glygen, Columbia, MD) and microplate porous graphitized carbon (PGC) SPE (Thermo Fisher Scientific, Waltham, MA) as follows. Between all additions, microplates were centrifuged at 1300 rpm, 20°C, for 1 min (Allegra X-30R Centrifuge, Beckman Coulter, Brea, CA). C18 microplates were conditioned with 200 µL acetonitrile (x 3), and equilibrated with 200 µL water (x 3). Samples (200 µL) were loaded atop of a fresh collection plate. The flowthrough and subsequent washes (3 x of 200 µL water) were collected for PGC SPE.

PGC microplates were conditioned with 200 µL 80/20 acetonitrile/water containing 0.1% TFA (x 5). Wells were equilibrated with 200 µL water (x 4). The collected flowthrough and washes from C18 SPE were loaded in 200 µL aliquots, centrifuging between additions. Wells were washed

with 200 µL water (x 6). A fresh collection plate was used to collect the eluent (3 x 200 µL 40/60 acetonitrile/water containing 0.1% TFA). Eluents were transferred to Eppendorf tubes and dried by centrifugal evaporation (Genevac miVac, Ipswich, United Kingdom). Samples were reconstituted in MilliQ water for nanoLC-QToF oligosaccharide analysis.

LC-MS/MS analysis was performed with an Agilent 6520 Accurate-Mass Q-TOF LC/MS instrument equipped with a Chip Cube coupled to an Agilent 1200 Series high performance liquid chromatography interface (Agilent Technologies, Santa Clara, CA) as described by Huang et al (2022). Data was annotated by Glyconote (https://github.com/MingqiLiu/GlycoNote) and manually verified using Agilent Masshunter Qualitative Analysis (B.07.00, Agilent Technologies). The data was manually searched for neutral loss of fucose and neutral loss of sulfate. Peaks were manually integrated using Agilent Masshunter Profinder (B.08.00, Agilent Technologies). Microsoft Excel was used for peak area quantitation and GraphPad Prism (ver. 9.4.0, GraphPad Software, LLC) was used to generate heatmaps from peak area data.

Monosaccharide analysis of algae extracts

Due to a lack of commercially available enzymes to aid in the study of the polysaccharide fucoidan, fucose (the primary monosaccharide in fucoidan) was used as a surrogate measure of fucoidan. Other less abundant monosaccharides (galactose, glucose, and mannose) making up less than 10% of fucoidan were also measured (Filote et al., 2021). These four monosaccharides are some of the most abundant monosaccharides that have been reported for fucoidan (Wang et al., 2020).

Method development was required to optimize hydrolysis time and instrument analysis. Published studies of fucoidan using trifluoroacetic acid hydrolysis consisted of temperatures ranging between 100-121°C and times ranging between 1 hour to 12 hours (Koh et al. 2019;

Lee et al., 2006; Zhang et al., 2012). To optimize for hydrolysis time, temperature was kept at a constant of 100 °C, and hydrolysis times of 2, 4, and 6 hours were tested. The hydrolysis time resulting in the highest concentration of the two most abundant monosaccharides (fucose and galactose) were selected, optimizing for maximum release of monosaccharides while minimizing monosaccharide degradation.

All selected extracts and enzyme blanks underwent acid hydrolysis in duplicate for monosaccharide analysis. Polysaccharides were precipitated from extracts by adding 3 volumes of cold ethanol. After centrifugation (14,000 x g, 4°C, 30 min) and drying, a pellet containing primarily polysaccharides was obtained. Polysaccharides were then hydrolyzed into constituent monosaccharides by 4 M trifluoroacetic acid (MilliporeSigma, St. Louis, MO) hydrolysis at 100 °C for 2 h with gentle stirring. Samples were briefly cooled, diluted with water to reach 0.5% trifluoroacetic acid, and acid was removed by centrifugal drying (Genevac miVac, Ipswich, United Kingdom). Hydrolysates were reconstituted in water then filtered by 0.2 µm polyethersulfone syringe filter (Pall Life Sciences, Port Washington, NY) for high-performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD) analysis. Samples were diluted appropriately to ensure monosaccharide quantification was within the linear range of the method.

To quantify monosaccharides, a Dionex HPAEC-PAD system was used to separate and detect fucose, galactose, glucose, and mannose. Analysis was performed on a Dionex ICS 5000+ with a detector consisting of an electrochemical cell with a disposable gold working electrode and a pH-Ag/AgCl reference electrode (Thermo Fisher Scientific, Waltham, MA). Samples were injected onto a Dionex CarboPac PA20 BioLC column (3 x 150 mm, Thermo Fisher Scientific) equipped with a PA20 guard column (3 x 30 mm, Thermo Fisher Scientific) at a flow rate of 0.5 mL min^{-1}. An isocratic chromatography method was developed and validated to ensure

reproducible results. Chromatographic separation was performed with an isocratic method consisting of 92% mobile phase A/8% mobile phase B (water (A) and 20 mM sodium hydroxide in water (B)) for 20 minutes. Commercial standards of the four monosaccharides were mixed and appropriately diluted to create a calibration curve. A linear range of 0.0001 - 0.01 mg mL^{-1} was established.

A calibration curve containing a mix of all four monosaccharide standards was used to quantify monosaccharide concentration in the extracts. Initial testing of monosaccharides in AEP samples showed that fucose and galactose were 1 to 2 magnitudes higher in concentration than glucose and mannose, consistent with previously reported findings (Wang et al., 2020). To ensure all monosaccharides were in linear range for quantification, two separate dilutions (1:2 and 1:100) were prepared and analyzed to ensure measurements were within the linear range of the calibration curve. Additionally, all samples were analyzed in duplicate to assess sample preparation and instrument reproducibility.

Results

Selected extraction conditions for functional compounds

Extractions were optimized for solid-to-liquid ratio, extraction temperature, time, pH, and enzyme concentration. Results from rapid assays, like colorimetric assays, for all extraction conditions were used to select specific extracts for further analysis (Gaiero et al., 2023). The selected extractions and their respective conditions are presented in Table 4.1. In total, nine extracts were selected for further protein and carbohydrate (oligosaccharide and monosaccharide) analysis.

Table 4.1. Nine selected extraction conditions for protein and carbohydrate analysis. The extraction method, pH (or enzyme addition, if relevant), temperature, and time are reported. All extractions maintained a solid-to-liquid ratio of 1:30.

Extraction process	pH (enzyme)	Extraction temperature (°C)	Time
Aqueous (AEP)	4	60	6 hr
Aqueous	7	60	6 hr
Aqueous	10	60	6 hr
Microwave-assisted aqueous	4	70	15 min
Microwave-assisted aqueous	7	70	15 min
Microwave-assisted aqueous	10	70	15 min
Microwave-assisted aqueous	pH 4 → pH 7	70	30 min
Microwave-enzyme aqueous	pH 4 → pH 7 (carbohydrase followed by protease)	60	30 min
Enzyme assisted (EAEP)	pH 4 → pH 7 (carbohydrase followed by protease)	60	4 hr

Proteomics results

Tentative proteomics results including a description and average mass of the identified protein are reported in Table 4.2. This analysis was not quantitative, thus results are reported as the presence or absence of a given protein for each AEP sample. Some key functions of the identified proteins are fixing atmospheric carbon dioxide (ribulose bisphosphate carboxylase) (Bathellier et al., 2020; Zhan et al., 2018) or photosynthesis (fucoxanthin-chlorophyll a-c binding, photosystem I reaction center subunit II) (Jordan et al., 2001; Nagao et al., 2014).

Table 4.2: Proteomics results of identified proteins and average protein mass (Da) for aqueous extraction (AEP) at pH 4, 7, and 10.

Protein Description	Avg. Mass (Da)	AEP pH 4	AEP pH 7	AEP pH 10
Ribulose bisphosphate carboxylase	53984	✓	✓	✓
ATP synthase subunit alpha	55210	✓	✓	✓
Multifunctional fusion protein	15840	✓		
ATP synthase subunit beta	52236	✓	✓	
Fucoxanthin-chlorophyll a-c binding protein	22730	✓	✓	
Ribosomal protein	9168	✓		
Photosystem I reaction center subunit II	15009	✓		
Ribulose bisphosphate carboxylase	50824	✓	✓	

These preliminary results, showing a decrease in the number of identified proteins as pH increases, appear to contradict protein content data gathered through combustion methods (Gaiero et al., 2023). It is likely that protein precipitation as described in the proteomics method section caused coprecipitation of high amounts of interfering compounds, namely polysaccharides (Nagai et al., 2008). Following the generation of this proteomics data, it was evident that significantly more method development would be required. Thus, proteomics analysis was not done for the remaining extraction conditions, and analytical efforts were instead focused on more salient monosaccharide and oligosaccharide analyses.

Glycoprofiling results

Glycomics was used to identify the composition of oligosaccharides released from giant kelp using various green extraction methods including aqueous, microwave-assisted, microwave-enzyme, and enzyme assisted extractions. A novel oligosaccharide library for giant kelp was generated through manual inspection of the data. Possible monosaccharides included hexoses (Hex), N-acetyl hexosamines (HexNAc), pentoses (Pent), deoxyhexose (dHex), and acidic

sugars (HexA). Additionally, features were manually searched for loss of dHex as would be expected for fucose-containing oligosaccharides stemming from fucoidan. All features in the giant kelp oligosaccharide library, including unknowns with demonstrated loss of dHex, are reported in Supplementary Table S4.1.

The number of identified oligosaccharides for each extraction condition and enzyme blank is reported in Figure 4.2. Microwave-assisted extracts, which only required 15 min of total processing time, had a comparable number of oligosaccharides identified to AEP extracts, which were extracted for 6 hours. Compared to microwave-assisted extraction, AEP had about the same number of identified oligosaccharides at neutral pH and a slightly higher number at basic pH. However, at acidic conditions, less oligosaccharides were identified in AEP vs. microwave-assisted extraction samples (34 compared to 39, respectively). This data demonstrates that the use of microwave-assisted extraction allows for a reduction in processing time from 6 hours to 15 minutes with minimal drawbacks.

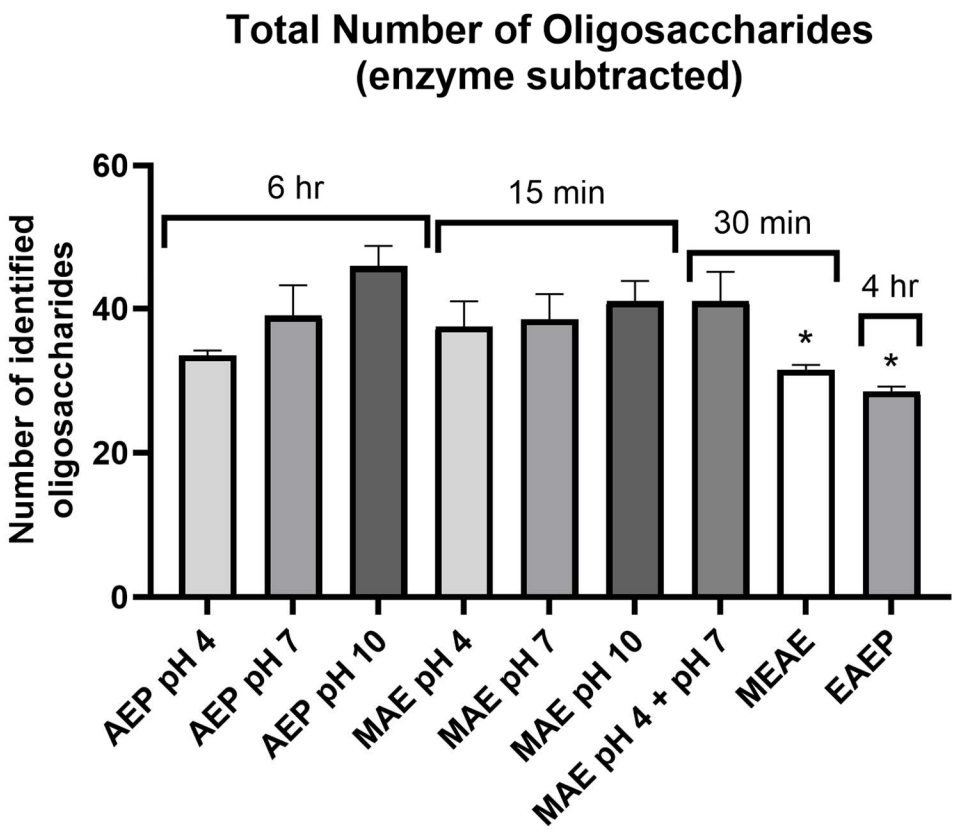


Figure 4.2. Total number of identified oligosaccharides for all selected extraction conditions (AEP = aqueous extraction, MAE = microwave-assisted extraction, MEAE = microwave-enzyme extraction, EAEP = enzyme-assisted extraction). Reported values have been subtracted by enzyme contributions. Labels above bars indicate the processing time for each method. The number of oligosaccharides identified in enzymes utilized in enzyme-assisted extractions are reported in the right two columns. *Values are likely underreported as a conservative estimate.

Further, the number of oligosaccharides identified with microwave-assisted extraction samples across all pH ranged from 39 to 44, while AEP samples across all pH ranged from 34 to 47. This seems to indicate that changes in pH make less of a difference for microwave-assisted extraction compared to AEP. Microwave-assisted extraction may be a more robust extraction process compared to traditional AEP, an important consideration especially for high throughput processing operations.

For AEP and microwave-assisted extraction samples, increasing pH resulted in higher numbers of identified oligosaccharides. For extractions involving enzymes (microwave-enzyme and EAEP), the enzymes used contained substantial amounts of oligosaccharides and required subtraction from the extracts. Given that the enzymes were expected to be carbohydrate-free, the abundance of carbohydrates with compositions corresponding to oligosaccharides in the enzyme preparations was surprising. Conservative estimates of the remaining compounds after enzyme subtraction are reported in Figure 4.2 (indicated with *). Without an experimental enzyme blank, enzyme contributions could not be quantitatively subtracted from the total amount detected. Thus, compounds were fully removed from enzyme-assisted extractions if they were also detected in the enzyme blank. As such, true values of the number of identified oligosaccharides for microwave-enzyme and EAEP samples are likely higher than reported.

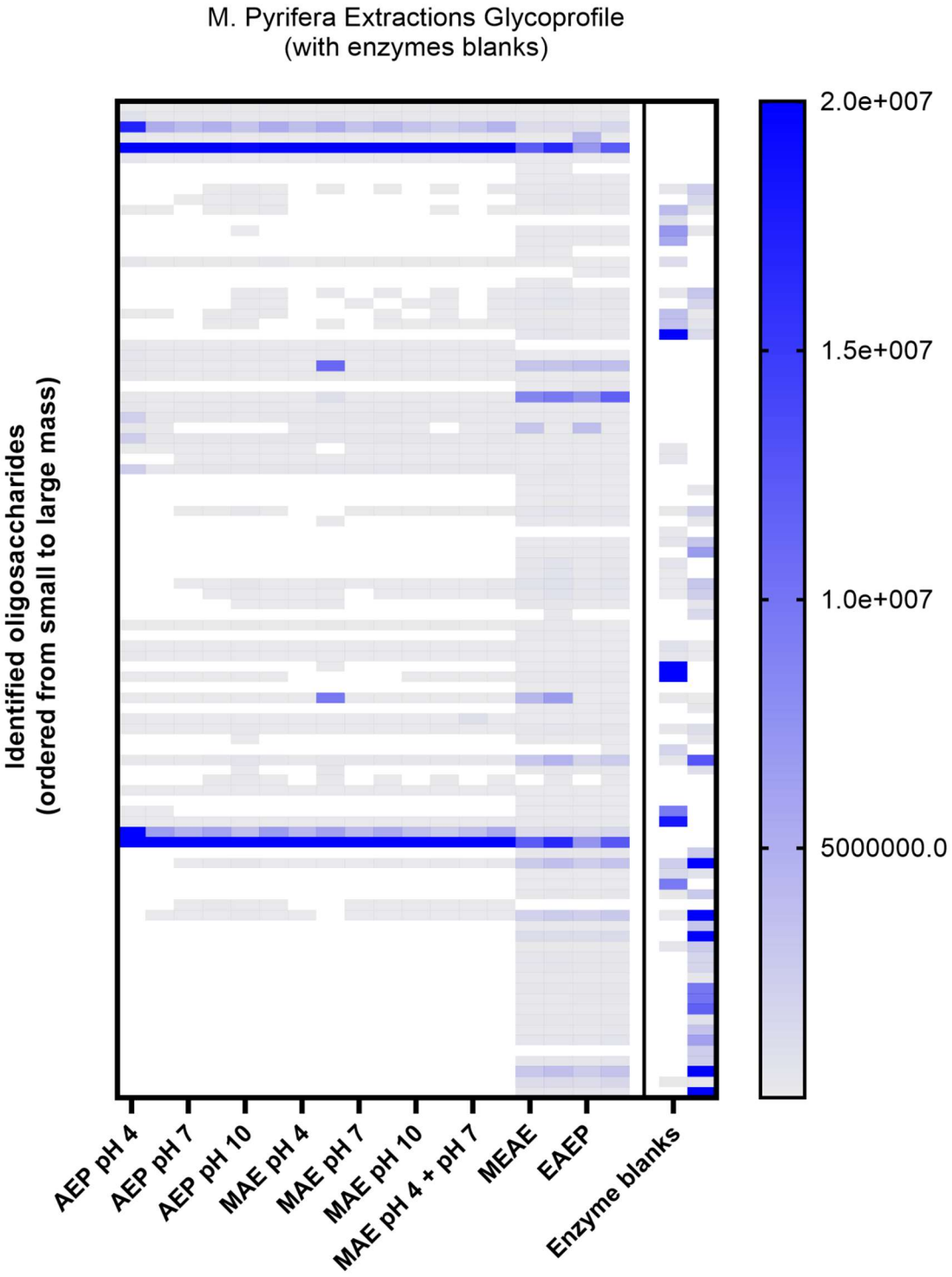
Figures 4.3A-B are heatmaps generated to visualize peak area data for all oligosaccharides and extractions. Individual oligosaccharides are reported on each row of the heatmap, and the four-digit codes along the y-axis indicate the number of monosaccharides making up the oligosaccharide (Hex_HexNAc_dHex_Pent_HexA). All 97 oligosaccharides identified by nanoLC-QToF, including those identified in the enzymes used in the extraction, are shown in Figure 4.3A. Figure 4.3B shows the identified oligosaccharides remaining after subtracting contributions made from extraction enzymes. The oligosaccharides identified solely in the enzyme blanks were plentiful and falsely inflated the total number of identified oligosaccharides stemming from giant kelp. These data demonstrate the importance of including experimental enzyme blanks in quantification and profiling studies.

Of the 61 oligosaccharides that remain after enzyme blank subtraction (Fig. 4.3B), 16 were identified as potentially containing fucose, but full identification of the oligosaccharide was not possible due to a lack of or obscure fragmentation data. Their neutral masses are reported

along the y-axis in Figure 4.3B. Among all oligosaccharides, two structures containing fucose but with unknown identities stand out. A compound with neutral mass of 473.223 Da and the dimerized form of this unknown (neutral mass 946.446 Da) were detected in much greater abundance compared to all other oligosaccharides. Other abundant oligosaccharides include 0_0_3_0_0 (an oligosaccharide consisting of 3 fucose units, likely a fucoidan fragment) and dimerized 0_0_3_0_0, which were detected among all extraction conditions with comparable abundances. Extractions using enzymes (microwave-enzyme and EAEP) also featured two fucose-containing but unknown compounds in relatively large abundance (neutral masses 512.214 Da and 527.231 Da). Due to the novelty of this research, the scientific literature does not yet report similar findings so results cannot be compared with published findings.

It is expected that fucoidan would contain primarily sulfated oligosaccharides with many fucose subunits (Haroun-Bouhedja et al., 2000), however none were detected through the methods described here. While fragmentation data did show many abundant features with neutral loss(es) of sulfate, the sulfate groups prevented the production of comprehensive fragmentation data that can be used to deduce the structure of the parent molecule (Filote et al., 2021; Pérez et al., 2016; Shi et al., 2012). Therefore, it is likely that all features reported in Figure 4.3B as “containing fucose” are unsulfated oligosaccharides.

A)


M. Pyrifera Extractions Glycoprofile
(with enzymes blanks)
Identified oligosaccharides
(ordered from small to large mass)
2.0e+007
1.5e+007
1.0e+007
5000000.0
AEP pH 4
AEP pH 7
AEP pH 10
MAE pH 4
MAE pH 7
MAE pH 10
MAE pH 4 + pH 7
MEAE
EAEP
Enzyme blanks

B)

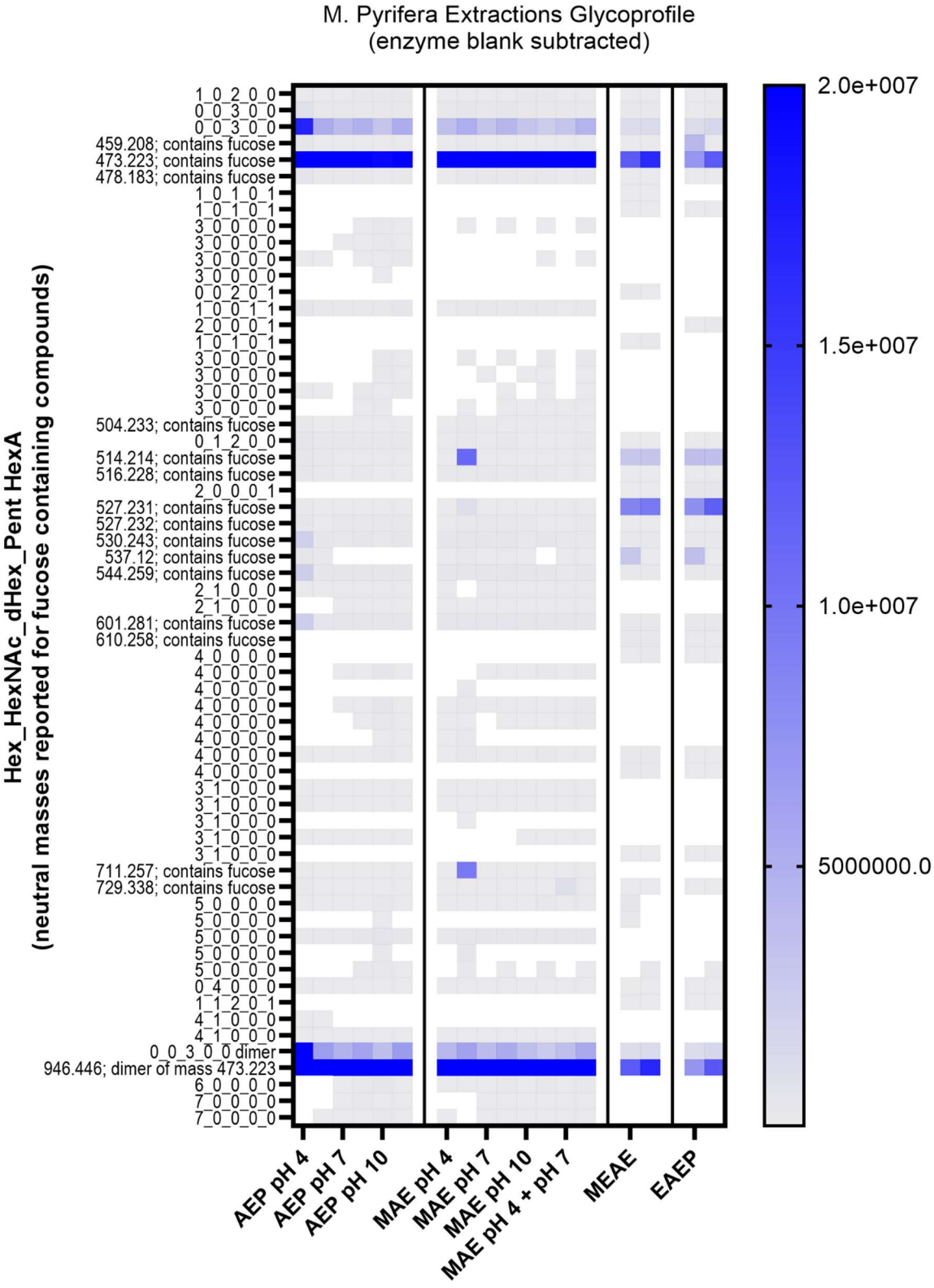

M. Pyrifera Extractions Glycoprofile
(enzyme blank subtracted)
Hex_HexNAc_dHex_Pent HexA
(neutral masses reported for fucose containing compounds)
1_0_2_0_0
0_0_3_0_0
0_0_3_0_0
459.208; contains fucose
473.223; contains fucose
478.183; contains fucose
1_0_1_0_1
1_0_1_0_1
3_0_0_0_0
3_0_0_0_0
3_0_0_0_0
3_0_0_0_0
0_0_2_0_1
1_0_0_1_1
2_0_0_0_1
1_0_1_0_1
3_0_0_0_0
3_0_0_0_0
3_0_0_0_0
3_0_0_0_0
504.233; contains fucose
0_1_2_0_0
514.214; contains fucose
516.228; contains fucose
2_0_0_0_1
527.231; contains fucose
527.232; contains fucose
530.243; contains fucose
537.12; contains fucose
544.259; contains fucose
2_1_0_0_0
2_1_0_0_0
601.281; contains fucose
610.258; contains fucose
4_0_0_0_0
4_0_0_0_0
4_0_0_0_0
4_0_0_0_0
4_0_0_0_0
4_0_0_0_0
4_0_0_0_0
4_0_0_0_0
3_1_0_0_0
3_1_0_0_0
3_1_0_0_0
3_1_0_0_0
3_1_0_0_0
711.257; contains fucose
729.338; contains fucose
5_0_0_0_0
5_0_0_0_0
5_0_0_0_0
5_0_0_0_0
5_0_0_0_0
0_4_0_0_0
1_1_2_0_1
4_1_0_0_0
4_1_0_0_0
0_0_3_0_0 dimer
946.446; dimer of mass 473.223
6_0_0_0_0
7_0_0_0_0
7_0_0_0_0
AEP pH 4
AEP pH 7
AEP pH 10
MAE pH 4
MAE pH 7
MAE pH 10
MAE pH 4 + pH 7
MEAE
EAEP
2.0e+007
1.5e+007
1.0e+007
5000000.0

Figure 4.3 Glycoprofile results showing all identified oligosaccharides (A) and results after enzyme subtraction (B). Glycoprofile of all identified oligosaccharides (A) are ordered from small to large masses (top to bottom) for selected extraction conditions (AEP = aqueous extraction, MAE = microwave-assisted extraction, MEAE = microwave-enzyme extraction, EAEP = enzyme-assisted extraction).and enzyme blanks. Identified oligosaccharides following enzyme blank subtraction for all selected extraction conditions (B) are reported by a 5-digit code, corresponding to the number of constituent monosaccharides (Hex_HexNAc_dHex_Pent_HexA) as shown along the y-axis. Unknown oligosaccharides containing fucose (dHex) are reported by their neutral mass.

Monosaccharide quantification

Method development was required to optimize hydrolysis time and instrument analysis. The two most abundant monosaccharides, fucose and galactose, and the extraction condition with the consistently highest monosaccharide concentrations, aqueous extraction at pH 4 (AEP pH 4) were used for hydrolysis time optimization. It was found that with 4 and 6 h of hydrolysis, monosaccharide content was much lower than that at 2 h of hydrolysis (Figure 4.4A-B). It is likely that this decrease in monosaccharide concentrations when hydrolysis times exceeded 2 h was due to degradation of the released monosaccharides. Two hours was chosen as the final hydrolysis time for all extracts to minimize monosaccharide degradation. In future work, hydrolysis times lower than 2 hours should be tested to find the optimal hydrolysis time for adequately hydrolyzing polysaccharides while minimizing monosaccharide degradation.

A) B)

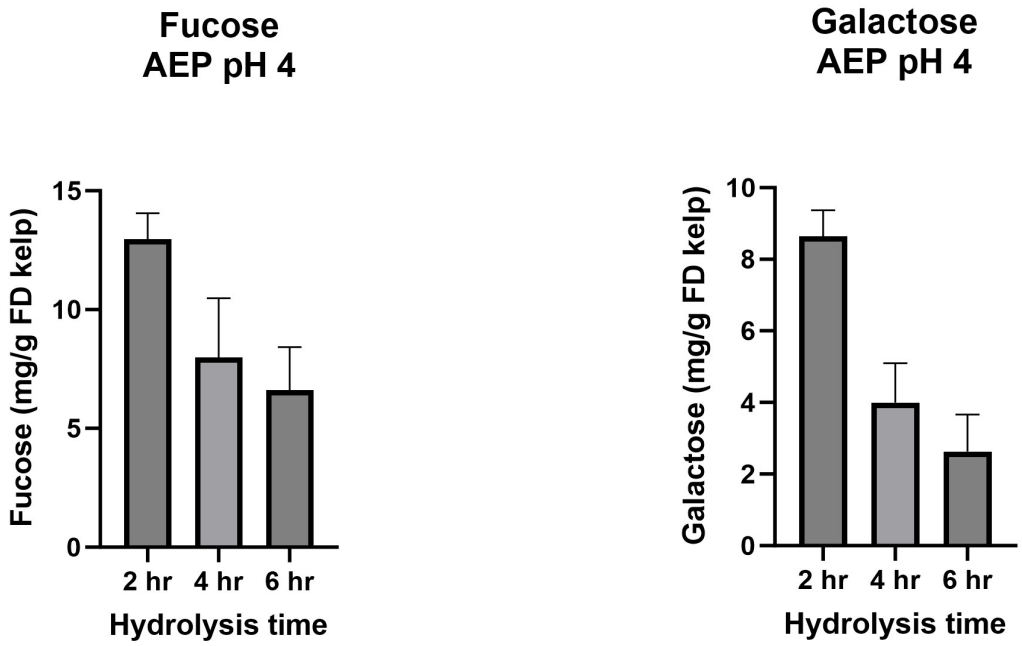


Figure 4.4. Monosaccharide content of fucose (A) and galactose (B) with varying hydrolysis times with aqueous extraction at pH 4 (AEP pH 4), reported in mg/g freeze dried kelp.

Monosaccharide quantification data for four monosaccharides, fucose, galactose, glucose, and mannose, are shown in Figure 4.5A-E. The highest concentration of fucose, used as surrogate measurement of the polysaccharide of interest, fucoidan, was 12.97 mg per g freeze dried kelp, which was associated with the AEP pH 4 extraction. For the majority of the measured monosaccharides, acidic conditions with aqueous or microwave-assisted extractions resulted in higher concentrations compared to neutral or basic conditions. These findings are reasonable, given that acidic conditions are typically used to hydrolyze polysaccharides and release monosaccharides (Koh et al., 2019; Lee et al., 2006; Zhang et al., 2012). The one deviation from this trend was mannose with aqueous extraction, which showed highest levels at pH 10, albeit its concentration was much lower than fucose and galactose.

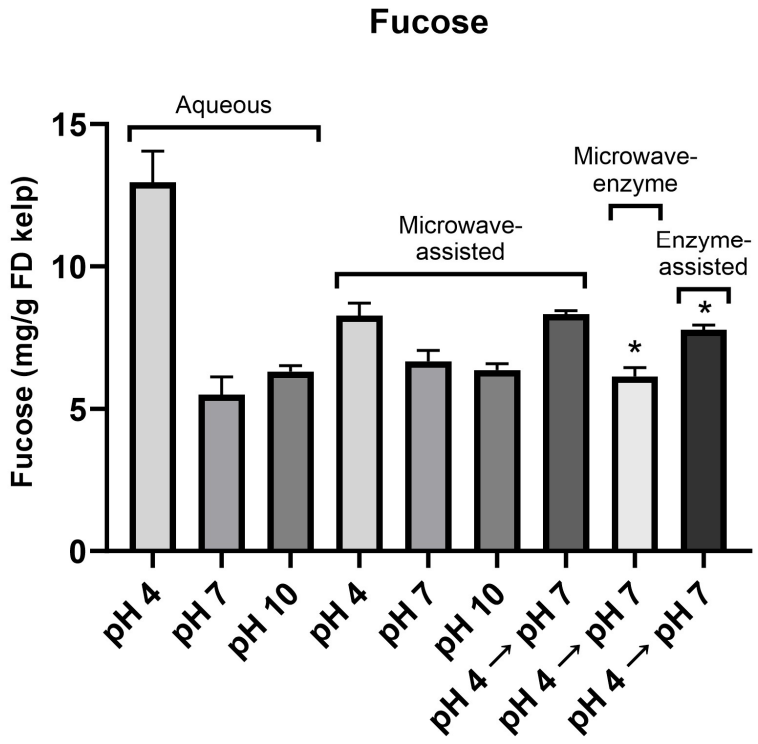
The use of enzymes in some extractions (enzyme-assisted and microwave-enzyme) created a confounding factor. The carbohydrase used had measurable levels of galactose, mannose, and xylose at 0.157, 0.050, and 0.131 mg mL^{-1}, respectively. The protease used had measurable levels of fucose, mannose, and glucose at 0.106, 23.574, and 0.166 mg mL^{-1}, respectively. These enzyme contributions were subtracted from the enzyme-assisted and microwave-enzyme measurements for their respective monosaccharides. However, for mannose, the enzyme contribution was so large compared to levels in native giant kelp that the enzyme contribution could not be accurately subtracted. As such, Figure 4.5E shows mannose levels for extractions with enzyme-assisted and microwave-enzyme extractions removed.

These data show that, with the exception of aqueous extraction at pH 4, there are only slight variations in monosaccharide content among all extraction conditions. These findings demonstrate that at neutral and basic conditions, 15 min of microwave-assisted extraction is roughly equivalent to 6 h of AEP. The use of microwave processing results in a significant decrease in processing time compared to aqueous extraction alone for roughly equivalent results. This may have valuable implications for commercial extraction of these compounds where processing time is an important consideration affecting the profitability of macroalgae processing.

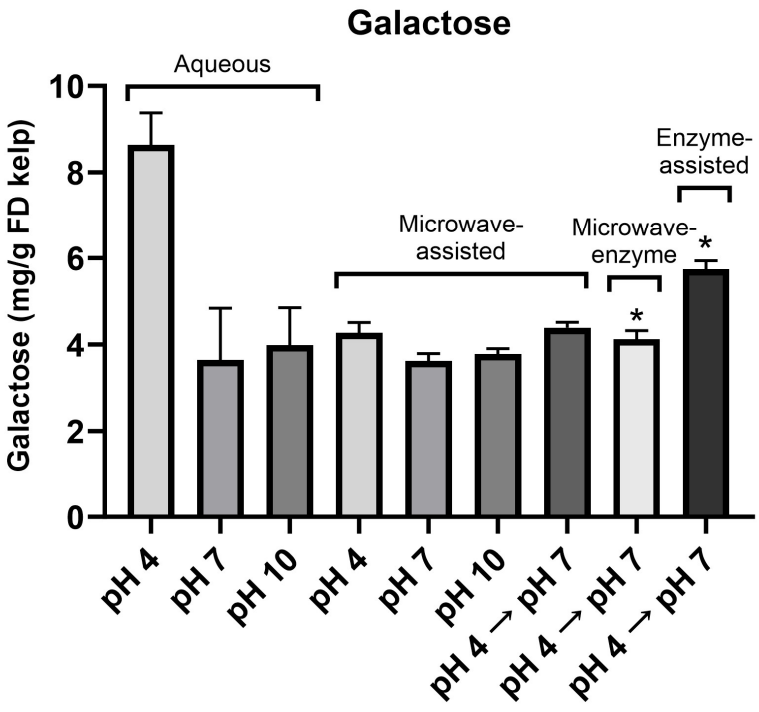
Additionally, monosaccharide concentrations of microwave-assisted extracts are fairly consistent at all tested pH, unlike AEP which had much lower concentrations at neutral and basic conditions compared to acidic conditions. The lack of pH dependence for microwave-assisted extraction methods may be one of the advantages of microwave-assisted extraction methods compared to AEP. This consistent extraction of polysaccharides along a wide range of pH for microwave-assisted extraction allows for easier extraction optimization for other targets of interest which may be more pH-dependent, like proteins. The flexibility of pH extraction

conditions allowed by microwave-assisted extraction methods increases the commercial feasibility of macroalgae processing at industrial scale.

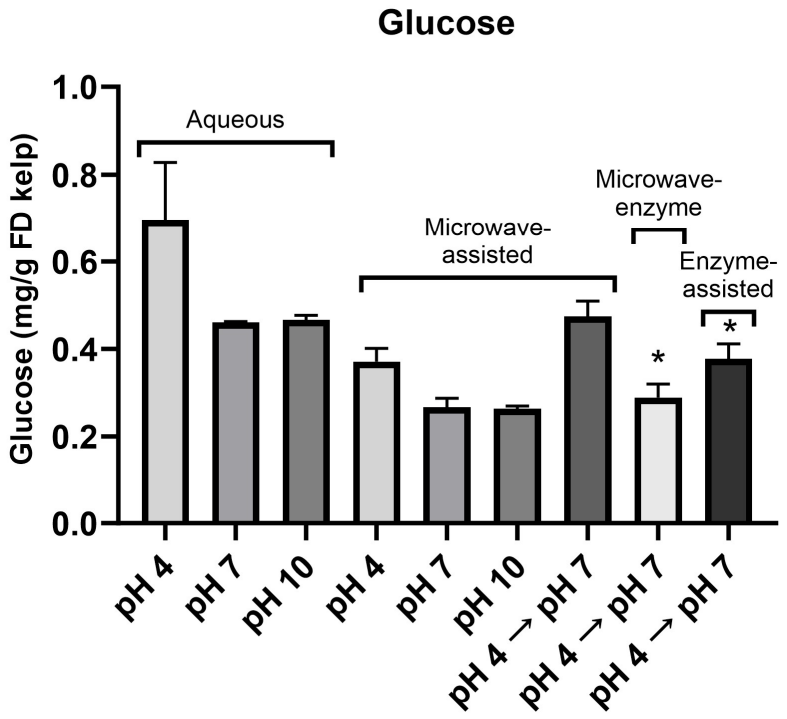
A)


Fucose
Aqueous
Microwave-assisted
Microwave-enzyme
Enzyme-assisted
Fucose (mg/g FD kelp)
15
10
5
0
*
*
pH 4
pH 7
pH 10
pH 4
pH 7
pH 10
pH 4 → pH 7
pH 4 → pH 7
pH 4 → pH 7

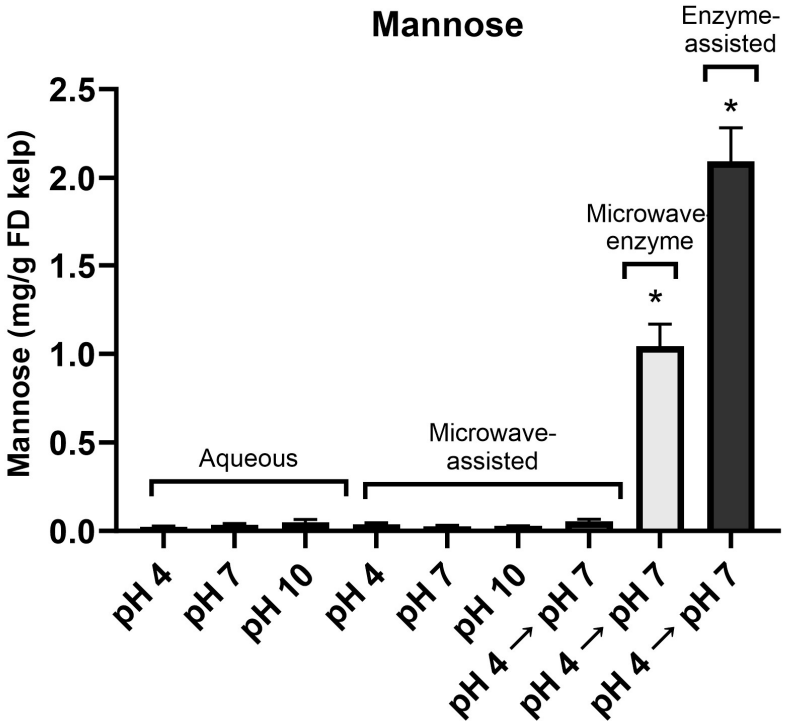
B)


Galactose
Aqueous
Microwave-assisted
Microwave-enzyme
Enzyme-assisted
Galactose (mg/g FD kelp)
10
8
6
4
2
0
*
*
pH 4
pH 7
pH 10
pH 4
pH 7
pH 10
pH 4 → pH 7
pH 4 → pH 7
pH 4 → pH 7

C)


Glucose
Glucose (mg/g FD kelp)
1.0
0.8
0.6
0.4
0.2
0.0
Aqueous
Microwave-assisted
Microwave-enzyme
Enzyme-assisted
*
*
pH 4
pH 7
pH 10
pH 4
pH 7
pH 10
pH 4 → pH 7
pH 4 → pH 7
pH 4 → pH 7

D)


Mannose
Mannose (mg/g FD kelp)
2.5
2.0
1.5
1.0
0.5
0.0
Aqueous
Microwave-assisted
Microwave enzyme
Enzyme-assisted
*
*
pH 4
pH 7
pH 10
pH 4
pH 7
pH 10
pH 4 → pH 7
pH 4 → pH 7
pH 4 → pH 7

E)

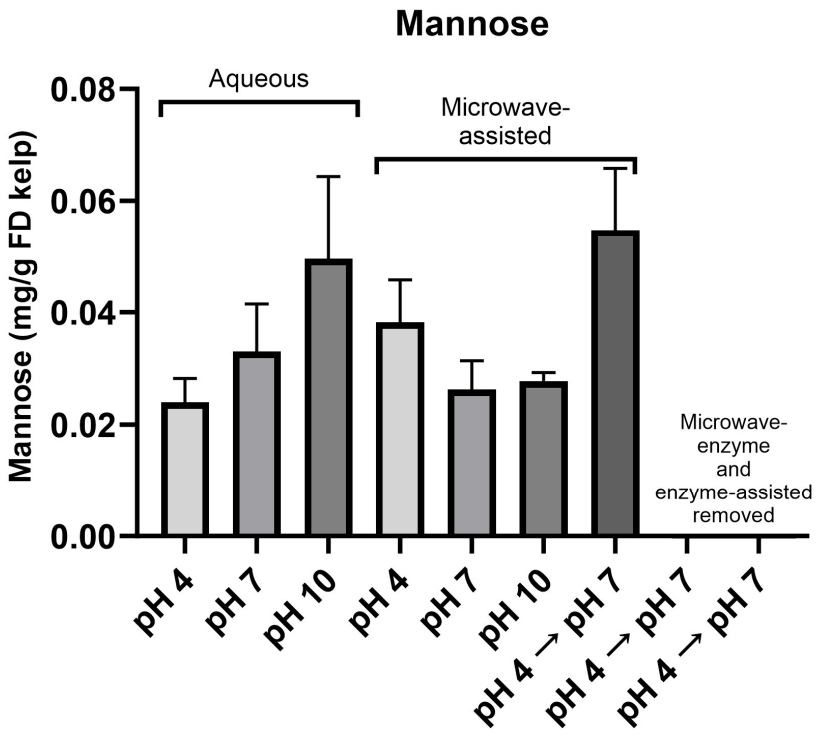


Figure 4.5. Monosaccharide concentrations of fucose (A), galactose (B), glucose (C), and mannose (D) in mg/g FD kelp. Labels above bar indicate the method of extraction used. For microwave-enzyme and enzyme-assisted extractions, enzyme contributions were subtracted (indicated with *). Figure 4.4E reports mannose concentrations without enzyme-containing samples, because enzymes used in extractions contained significantly more mannose than from native giant kelp, and enzyme contributions were unable to be accurately subtracted.

Bioactivity results (antidiabetic activity)

α-glucosidase inhibition is one method of managing diabetes. The α-glucosidase inhibitory activity of the extracts was evaluated (Gaiero et al., 2023), and the results are presented in Figure 4.6 as percent inhibition. The results show comparable inhibition levels across the tested conditions. To gain further insights into the differences between extraction conditions, additional dilutions are required in the future. Nonetheless, these results provide valuable information when compared to the commercially available drug acarbose. Interestingly, all of the extracts exhibited inhibition levels exceeding 90% at a concentration of 0.5 mg freeze-dried extract mL^{-1}, while a 4 mg mL^{-1} solution of acarbose resulted in approximately 80% inhibition. For comparison, the recommended therapeutic dose of acarbose for diabetes management is 25 mg, three times a day (Mayo Clinic, 2023). These findings demonstrate the potential therapeutic value of these extracts in the management of diabetes.

The importance of these findings holds great potential, particularly in the context of microwave-assisted extraction which took only 15 minutes and yielded superior results compared to existing market drugs. Other studies on various macroalgae have also reported α-glucosidase inhibition activity surpassing that of acarbose (Yuan et al., 2018). The antidiabetic effects have been attributed to algal polysaccharides, such as fucoidans (Jia et al., 2022; Yuan et al., 2018).

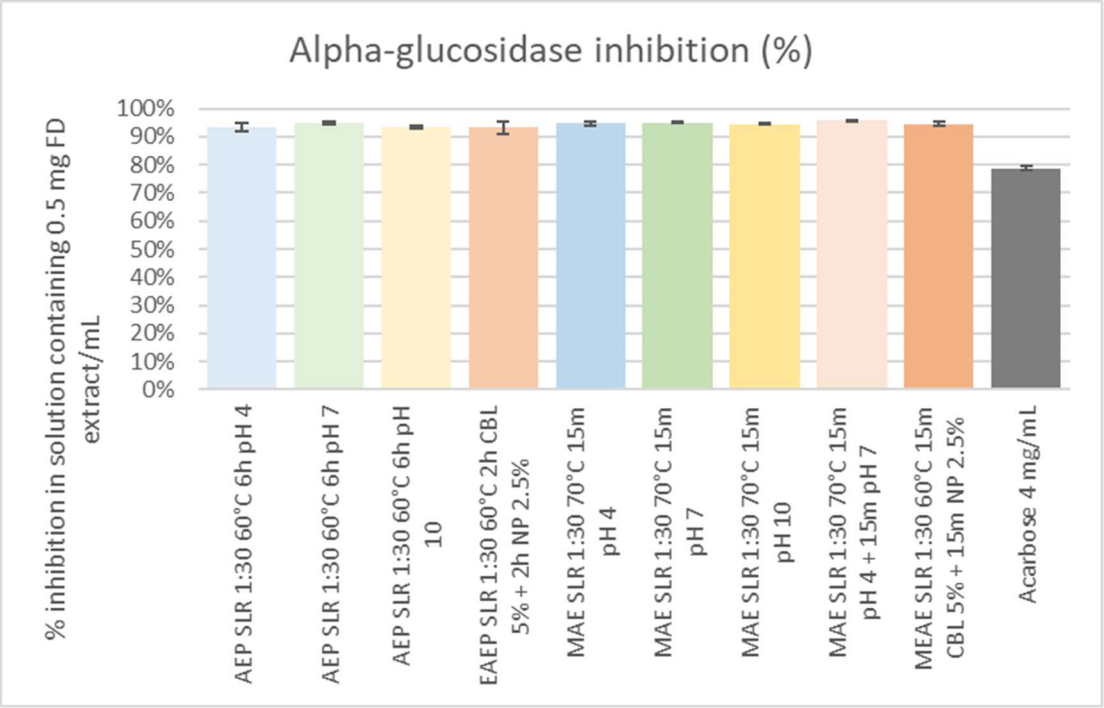


Figure 4.6. α-glucosidase inhibition activity (%) of selected extraction conditions (AEP = aqueous extraction, EAEP = enzyme-assisted extraction, MAE = microwave-assisted extraction, MEAE = microwave-enzyme extraction) compared to market drug Acarbose (d).

Discussion

Algae proteins

Proteomics was employed to identify proteins released using different extraction methods. The findings from aqueous extractions revealed the identification of nine proteins, with acidic conditions yielding the highest number of identified proteins. However, these data were inconsistent with protein analysis results obtained through alternative methods such as combustion techniques. Further, the total number of identified proteins was considerably lower than expected given the sample sizes used for analysis and the size of the protein database for giant kelp. These observations suggested potential issues related to proteomics sample preparation or analysis.

These challenges can likely be attributed to the relatively low concentration of proteins in giant kelp and the considerably higher polysaccharide content present. The low protein concentration in the extracts posed difficulties during preliminary data acquisition. Proteomics analysis requires sufficient concentrations of proteins for effective digestion and peptide detection. Typical colorimetric and fluorometric assays for protein quantification could not be used due to the dark color and viscous nature of the extracts. Alternatively, SDS-PAGE analysis can be utilized to visualize proteins and provide context for proteomics data. However, the extracts did not contain sufficient protein concentrations to yield discernible bands on the gel. Concentrating the extracts by centrifugal evaporation resulted in an increasingly viscous solution, preventing any present proteins from entering the gel.

This preliminary data gathering stage is crucial for ensuring reliable results in proteomics analysis due to two reasons. First, trypsin digestion requires a relatively high protein concentration of 1 mg/mL for effective digestion. Second, obtaining proteomics data is a time-intensive process, requiring 3-4 days of extensive sample preparation and 2 h of instrument analysis time per sample. Given the challenges encountered with obtaining preliminary data using SDS-PAGE and various protein quantification assays, proteomics analysis was forced to proceed without the typical data used to guide the sample preparation steps. Alternative steps were taken to obtain a surrogate measure of protein with the aim of ensuring adequate protein concentrations throughout sample preparation.

To estimate protein content, dried protein pellets obtained after protein precipitation were weighed as a surrogate measure. Protein amounts were estimated to be 10 times lower than the pellet weight to ensure adequate detection of peptides, and the final sample volume was minimized to maximize peptide concentration. Despite these adjustments, the samples remained too dilute to measure protein concentration by Qubit 3.0 Fluorometer (Thermo Fisher

Scientific, Waltham, MA), a common metric for adequate instrument response. Samples were analyzed by nanoLC-QToF with the maximum injection volume. Although adequate instrument response was not reached due to the dilute nature of the extracted proteins, some data was obtained (Table 4.2).

However, the obtained proteomics results revealed another challenge—the data did not align with the protein content data obtained through Dumas combustion analysis (Gaiero et al., 2023). When compared to the combustion data, the proteomics data revealed an inverse result. Dumas combustion shows that protein amounts increased with increased pH for aqueous extracts, whereas the proteomics data demonstrated a decrease in the number of identified proteins with increasing pH.

Further investigation of the literature suggested that the discrepancy may be due to co-precipitation of inferring compounds, namely polysaccharides (Nagai et al., 2008). In giant kelp, protein content is relatively low compared to carbohydrate content (6.5% protein vs. 63.4% carbohydrate) (Gaiero et al., 2023). The inability to separate small amounts of protein from large amounts of polysaccharides resulted in a general lack of efficiency for trypsin digestion and protein detection and identification. The literature offers some potential solutions to reducing this co-precipitation of proteins and polysaccharides (Contreras et al., 2008; Nagai et al., 2008; Wong et al., 2006), however none are universal, and all are dependent on the type of sample and intended analysis. Considering the limited timeline of the project and the overall low protein content compared to carbohydrates in giant kelp, no further method development was pursued, and research efforts shifted to focusing on the abundance of carbohydrates instead.

Nevertheless, despite the low protein content in giant kelp, this work would be valuable to revisit in the future. There is evidence of bioactive peptides isolated from various algae species that

exhibit antihypertensive, antioxidative, and antidiabetic effects (Admassu et al., 2018; Biparva, et al., 2023). Additionally, amino acid profiling of giant kelp reveals the presence of all essential amino acids, making it potentially valuable for food and nutraceutical applications (Gaiero et al., 2023). Although the novel nature of this work presented some challenges, these efforts helped to increase understanding of the impact of green extraction methods and unlock giant kelp's potential.

Algae carbohydrates

The carbohydrate fraction of giant kelp extracts was investigated by analyzing free oligosaccharides and the monosaccharides obtained through acid hydrolysis. Glycoprofiling showed minor differences between 6 h of aqueous extraction and 15 min of microwave-assisted extraction. Further, microwave-assisted extraction demonstrated less dependence on pH, thus it was more robust and offered additional optimization opportunities for other compounds, such as proteins. Similar findings were observed in the monosaccharide analysis. At neutral and basic pH, monosaccharide contents were comparable between 6 h aqueous extraction and 15 min microwave-assisted extraction. These results indicate the potential of microwave-assisted extraction as a viable green extraction technology for industrial applications. However, the analysis of carbohydrates in this study presented several significant challenges, including the lack of commercially available enzymes for studying polysaccharides of interest, LCMS analysis limitations, and contaminations from enzymes used in extraction.

One primary challenge in the polysaccharide analysis of giant kelp was the lack of commercially available enzymes that could be used to study large polysaccharides, specifically fucoidan. We were unable to find macroalgae specific enzymes, as this is still a fairly novel area and specific enzymes are not commercially available. Such enzymes are necessary to hydrolyze large polysaccharides into smaller oligosaccharides suitable for mass spectrometry analysis. While

quadrupole time-of-flight (Q-TOF) mass spectrometers (employed for oligosaccharide analysis for this project) can operate in wide mass ranges, sensitivity and resolving power decrease with increasing molecular weight, making accurate detection of large molecules challenging (El-Aneed, 2009). As a polysaccharide, fucoidan ranges in mass from 100 to 1600 kDA (Rioux, et al., 2007), greatly exceeding the mass range of the analytical instrument. Therefore, obtaining enzymes capable of breaking down polysaccharides in a predictable manner are crucial for mass spectrometry analysis. In theory, fucoidan could be characterized through fragmentation data and with knowledge of which specific linkages have been cleaved by an enzyme. These fragments could then be assembled to obtain a comprehensive understanding of branching, monosaccharide substituents, and degrees of sulfation. This approach would allow investigation into the diverse structures of fucoidan that are extracted by different extraction methods. Importantly, these variable structures, dictated by the amount of branching, degree of sulfation, and constituent monosaccharides, have been shown to affect the biological activity of fucoidan (Haroun-Bouhedja et al., 2000). This study, along with others, found high α-glucosidase inhibition activity in these giant kelp extracts, which have been attributed to polysaccharides such as fucoidan (Gaiero et al., 2023; Jia et al., 2022; Yuan et al., 2018). Access to enzymes capable of hydrolyzing large polysaccharides through predictable cleavage is necessary to fully characterize fucoidan by mass spectrometry. Without these enzymes, surrogate measures like fucose quantification were employed to estimate fucoidan levels.

In contrast to proteomics analysis, where challenges mainly arose during sample preparation due to low protein abundance, oligosaccharides were abundant and more easily isolated and prepared for instrument analysis. However, oligosaccharide analysis presented challenges during sample preparation for instrument analysis and data processing for determining oligosaccharide composition. The primary polysaccharide of interest, fucoidan, is expected to be 100-1600 kDa (Kadam et al., 2015). For LC-MS sample preparation, solid phase extraction

using PGC was employed. It is possible that PGC prevents the elution of larger oligosaccharides and polysaccharides, such as fucoidan, and only smaller oligosaccharides were obtained following PGC. Further, oligosaccharides and polysaccharides present ionization and fragmentation challenges. Intact versions of polysaccharides cannot be directly analyzed, so oligosaccharides are used as surrogate markers. Consequently, it can be difficult to determine which polysaccharide an oligosaccharide might have stemmed from, preventing the ability to make any confident conclusions about a particular polysaccharide. Lastly, the high degree of sulfation expected with fucoidan made elucidating structure from fragmentation data challenging. The highly labile bond of sulfate groups compared to other bonds (Shi et al., 2012) made the loss of neutral sulfate the dominant ion and limited the number of fragments that can be used as structural information. The data show that neutral loss of sulfate was detected in abundance, but fragmentation of the rest of the molecule was poor compared to loss of sulfate, preventing the identification of sulfate-containing oligosaccharides. To overcome this challenge, chemical modification or sulfate groups may be required so that neutral loss of sulfate is not the dominant fragment ion.

Another challenge arose from the enzymes used in enzyme-assisted extraction processes. Monosaccharide and oligosaccharide analysis revealed that the enzymes used for extractions contained easily detectable amounts of these compounds, falsely inflating some measurements and in many cases, introducing oligosaccharides that are wholly absent in native giant kelp. The contaminations from enzymes were substantial with 36 out of the 97 total identified oligosaccharides in giant kelp extracts attributed solely to the two enzymes (Figure 4.2A). These contaminants, likely fragments derived from the cell wall of the organisms used in the production of modern recombinant enzymes, hindered the ability to gain accurate quantitative oligosaccharide and monosaccharide data for any extractions involving enzymes. Similar enzyme contaminations have been reported in commercially available kits utilizing β-

glucosidase (Pharr and Dickinson, 1973). In future work, including enzyme blanks in both extraction and analysis is crucial, as these enzymes have shown to significantly contribute to the data. This is especially important considering it is possible that the composition of the enzymes may vary between batches or vendors. As such, purchasing from a consistent batch or vendor may be important considerations especially for comparative studies over long periods of time.

Conclusion

The application of proteomics to identify proteins released from giant kelp using different extraction methods provided valuable insights and highlighted challenges in sample preparation and analysis. The study revealed a lower-than-expected number of identified proteins, likely due to the relatively low concentration of proteins and the high polysaccharide content in giant kelp. The limitations encountered in preliminary data acquisition required alternative approaches, but despite the challenges, some proteomics data was obtained. However, the proteomics results did not align with protein data acquired through other means, suggesting potential issues related to co-precipitation of proteins and polysaccharides. Although the project's focus shifted to investigating the abundance of carbohydrates, the work remains valuable as it contributes to a deeper understanding of green extraction methods and the potential applications of giant kelp. Future research should address these encountered challenges, given the evidence of bioactive peptides and essential amino acids found in giant kelp, which hold promise for various applications in the food and nutraceutical industries.

The use of carbohydrate analysis to characterize oligosaccharides and quantify monosaccharides provided valuable insights in the development of green extraction methods for giant kelp. This study demonstrated the potential of microwave-assisted extraction as a green, robust method with optimization opportunities. When comparing monosaccharide contents and

glycoprofiles, 15 min microwave-assisted extraction was comparable to 6 h aqueous extraction. Extracts of giant kelp demonstrated strong bioactivity, even exceeding those of equivalent market drugs in the case of α-glucosidase inhibition. These findings are valuable for potential applications and industrial processing of giant kelp. However, challenges in carbohydrate analysis were encountered, particularly in the lack of commercially available enzymes for studying large polysaccharides like fucoidan and difficulties gaining structural insight using mass spectrometry. Additionally, the study identified contaminations from extraction enzymes, emphasizing the importance of including enzyme blanks in future work to ensure accurate quantitative data. Overcoming these challenges is crucial for gaining a comprehensive understanding of the diverse structures of algal polysaccharides and their associated biological activities. This work is the first step in gaining a deeper understanding of the influence of green extraction methods and the numerous potential functional and biological benefits offered by giant kelp.

Acknowledgements

This research was supported by funding from AlgaePrize (U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy Bioenergy Technologies Office) and Bigelow Laboratory for Ocean Sciences.

Supplementary materials

Supplementary Table S4.1. Compositions, neutral mass, and retention time (minutes) of all oligosaccharides confirmed by tandem MS/MS in at least one of the giant kelp extractions. Features with demonstrated loss of fucose (dHex) are also reported as unknowns.

Hex_HexNAc_dHex_Pent_HexA	Mass	Retention time (min)	Hex	HexNAc	dHex	Pent	HexA
1_0_2_0_0	454.1686	13.8	1	0	2	0	0
0_0_3_0_0	456.1843	10.2	0	0	3	0	0
0_0_3_0_0	456.1843	14.5	0	0	3	0	0
0_0_3_0_0	456.1843	16.1	0	0	3	0	0
unknown; loss of fucose	459.2051	11.5					
unknown; loss of fucose	473.222	10.2					
unknown; loss of fucose	473.222	11.0					
unknown; loss of fucose	478.1799	11.2					
1_0_1_0_1	484.1428	16.9	1	0	1	0	1
1_0_1_0_1	484.1428	18.3	1	0	1	0	1
3_0_0_0_0	486.1584	5.4	3	0	0	0	0
3_0_0_0_0	486.1584	6.4	3	0	0	0	0
3_0_0_0_0	486.1584	9.5	3	0	0	0	0
3_0_0_0_0	486.1584	10.5	3	0	0	0	0

3_0_0_0_0	486.1584	12.4	3	0	0	0	0
3_0_0_0_0	486.1584	15.0	3	0	0	0	0
3_0_0_0_0	486.1584	17.5	3	0	0	0	0
3_0_0_0_0	486.1584	18.6	3	0	0	0	0
0_0_2_0_1	486.1585	25.3	0	0	2	0	1
1_0_0_1_1	488.1378	15.3	1	0	0	1	1
unknown; loss of fucose	489.2161	11.0					
2_0_0_0_1	500.1377	19.0	2	0	0	0	1
1_0_1_0_1	502.1534	16.9	1	0	1	0	1
3_0_0_0_0	504.169	5.5	3	0	0	0	0
3_0_0_0_0	504.169	6.5	3	0	0	0	0
3_0_0_0_0	504.169	10.5	3	0	0	0	0
3_0_0_0_0	504.169	12.5	3	0	0	0	0
3_0_0_0_0	504.169	17.2	3	0	0	0	0
3_0_0_0_0	504.169	18.4	3	0	0	0	0
unknown; loss of fucose	504.2313	10.0					
0_1_2_0_0	513.2058	15.2	0	1	2	0	0
unknown; loss of fucose	514.2017	15.1					

unknown; loss of fucose	516.228	10.5					
2_0_0_0_1	518.1483	19.0	2	0	0	0	1
unknown; loss of fucose	527.234	14.4					
unknown; loss of fucose	527.234	16.2					
0_0_0_4_0	528.1692	18.6	0	0	0	4	0
0_0_0_4_0	528.1692	23.7	0	0	0	4	0
unknown; loss of fucose	530.2425	11.7					
unknown; loss of fucose	531.2151	12.7					
unknown; loss of fucose	537.1181	16.3					
unknown; loss of fucose	539.219	26.2					
unknown; loss of fucose	544.2565	12.5					
2_1_0_0_0	545.1956	9.4	2	1	0	0	0
2_1_0_0_0	545.1956	11.3	2	1	0	0	0
0_0_4_0_0	584.2316	16.7	0	0	4	0	0
unknown; loss of fucose	601.2778	13.6					
unknown; loss of fucose	601.2799	13.2					
unknown; loss of fucose	610.2576	26.2					
unknown; loss of fucose	624.2376	24.8					

4_0_0_0_0	648.2112	6.5	4	0	0	0	0
4_0_0_0_0	648.2112	10.0	4	0	0	0	0
4_0_0_0_0	648.2112	10.6	4	0	0	0	0
4_0_0_0_0	648.2112	12.0	4	0	0	0	0
4_0_0_0_0	648.2112	15.2	4	0	0	0	0
4_0_0_0_0	648.2112	16.0	4	0	0	0	0
4_0_0_0_0	648.2112	25.5	4	0	0	0	0
1_0_2_0_1	648.2113	16.3	1	0	2	0	1
4_0_0_0_0	666.2218	5.5	4	0	0	0	0
4_0_0_0_0	666.2218	6.5	4	0	0	0	0
4_0_0_0_0	666.2218	9.8	4	0	0	0	0
4_0_0_0_0	666.2218	10.5	4	0	0	0	0
4_0_0_0_0	666.2218	12.2	4	0	0	0	0
4_0_0_0_0	666.2218	13.1	4	0	0	0	0
4_0_0_0_0	666.2218	15.0	4	0	0	0	0
4_0_0_0_0	666.2218	25.2	4	0	0	0	0
4_0_0_0_0	666.2218	25.7	4	0	0	0	0
3_0_0_0_1	680.2011	24.4	3	0	0	0	1

3_1_0_0_0	707.2484	11.0	3	1	0	0	0
3_1_0_0_0	707.2484	12.5	3	1	0	0	0
3_1_0_0_0	707.2484	13.7	3	1	0	0	0
3_1_0_0_0	707.2484	15.2	3	1	0	0	0
3_1_0_0_0	707.2484	16.5	3	1	0	0	0
3_1_0_0_0	707.2484	22.5	3	1	0	0	0
unknown; loss of fucose	711.2708	14.8					
unknown; loss of fucose	712.3085	19.6					
0_0_3_2_0	720.2689	18.9	0	0	3	2	0
unknown; loss of fucose	729.338	15.7					
5_0_0_0_0	810.264	12.0	5	0	0	0	0
5_0_0_0_0	810.264	13.4	5	0	0	0	0
5_0_0_0_0	810.264	17.0	5	0	0	0	0
5_0_0_0_0	828.2746	12.0	5	0	0	0	0
5_0_0_0_0	828.2746	13.4	5	0	0	0	0
5_0_0_0_0	828.2746	29.0	5	0	0	0	0
0_4_0_0_0	830.3282	15.6	0	4	0	0	0
4_0_0_0_1	842.2539	28.0	4	0	0	0	1

1_1_2_0_1	853.3063	15.2	1	1	2	0	1
1_1_2_0_1	853.3063	16.2	1	1	2	0	1
4_1_0_0_0	869.3012	10.7	4	1	0	0	0
4_1_0_0_0	869.3012	11.5	4	1	0	0	0
unknown; dimer of mass 474.2293	946.4471	10.2					
6_0_0_0_0	972.3168	13.4	6	0	0	0	0
6_0_0_0_0	990.3274	13.4	6	0	0	0	0
6_0_0_0_0	990.3274	14.4	6	0	0	0	0
6_0_0_0_0	990.3274	15.0	6	0	0	0	0
6_0_0_0_0	990.3274	15.5	6	0	0	0	0
5_0_0_0_1	1004.307	13.0	5	0	0	0	1
7_0_0_0_0	1134.37	14.5	7	0	0	0	0
7_0_0_0_0	1134.37	19.4	7	0	0	0	0
7_0_0_0_0	1152.38	14.4	7	0	0	0	0
7_0_0_0_0	1152.38	15.7	7	0	0	0	0
5_0_0_3_0	1206.282	14.8	5	0	0	3	0
8_0_0_0_0	1296.422	15.6	8	0	0	0	0
8_0_0_0_0	1314.433	15.6	8	0	0	0	0

8_0_0_0_0	1314.433	16.7	8	0	0	0	0
9_0_0_0_0	1458.475	17.0	9	0	0	0	0
9_0_0_0_0	1476.486	17.0	9	0	0	0	0
9_0_0_0_0	1476.486	17.5	9	0	0	0	0
10_0_0_0_0	1620.528	18.5	10	0	0	0	0
10_0_0_0_0	1638.539	18.6	10	0	0	0	0
10_0_0_0_0	1638.539	19.2	10	0	0	0	0
11_0_0_0_0	1800.591	20.4	11	0	0	0	0
12_0_0_0_0	1944.634	14.0	12	0	0	0	0
12_0_0_0_0	1962.623	21.9	12	0	0	0	0

References

Abraham, R. E., Su, P., Puri, M., Raston, C. L., and Zhang, W. (2019). Optimisation of biorefinery production of alginate, fucoidan and laminarin from brown seaweed Durvillaea potatorum. *Algal Research*, *38*, 101389. https://doi.org/10.1016/j.algal.2018.101389

Admassu, H., Gasmalla, M. A. A., Yang, R., and Zhao, W. (2018). Bioactive Peptides Derived from Seaweed Protein and Their Health Benefits: Antihypertensive, Antioxidant, and Antidiabetic Properties. *Journal of Food Science*, *83*(1), 6–16. https://doi.org/10.1111/1750-3841.14011

Alboofetileh, M., Rezaei, M., Tabarsa, M., Rittà, M., Donalisio, M., Mariatti, F., … Cravotto, G. (2019). Effect of different non-conventional extraction methods on the antibacterial and antiviral activity of fucoidans extracted from Nizamuddinia zanardinii. *International Journal of Biological Macromolecules*, *124*, 131–137. https://doi.org/10.1016/j.ijbiomac.2018.11.201

Alboofetileh, M., Rezaei, M., and Tabarsa, M. (2018). Enzyme-assisted extraction of Nizamuddinia zanardinii for the recovery of sulfated polysaccharides with anticancer and immune-enhancing activities. *Journal of Applied Phycology*, *31*(2), 1–12. https://doi.org/10.1007/s10811-018-1651-7

Andriamanantoanina, H., and Rinaudo, M. (2010). Characterization of the alginates from five madagascan brown algae. *Carbohydrate Polymers*, *82*(3), 555–560. https://doi.org/10.1016/j.carbpol.2010.05.002

Atashrazm, F., Lowenthal, R. M., Woods, G. M., Holloway, A. F., and Dickinson, J. L. (2015). Fucoidan and cancer: a multifunctional molecule with anti-tumor potential. *Marine Drugs*, *13*(4), 2327–2346. https://doi.org/10.3390/md13042327

Barbarino, E., and Lourenço, S. O. (2005). An evaluation of methods for extraction and quantification of protein from marine macro- and microalgae. *Journal of Applied Phycology*, *17*(5), 447–460. https://doi.org/10.1007/s10811-005-1641-4

Bathellier, C., Yu, L.-J., Farquhar, G. D., Coote, M. L., Lorimer, G. H., and Tcherkez, G. (2020). Ribulose 1,5-bisphosphate carboxylase/oxygenase activates O2 by electron transfer. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(39), 24234–24242. https://doi.org/10.1073/pnas.2008824117

Biparva, P., Mirzapour-Kouhdasht, A., Valizadeh, S., and Garcia-Vaquero, M. (2023). Advanced Processing of Giant Kelp (Macrocystis pyrifera) for Protein Extraction and Generation of Hydrolysates with Anti-Hypertensive and Antioxidant Activities In Vitro and the Thermal/Ionic Stability of These Compounds. *Antioxidants (Basel, Switzerland)*, *12*(3). https://doi.org/10.3390/antiox12030775

Bleakley, S., and Hayes, M. (2017). Algal proteins: extraction, application, and challenges concerning production. *Foods*, *6*(5). https://doi.org/10.3390/foods6050033

Bordoloi, A., and Goosen, N. (2020). Green and integrated processing approaches for the recovery of high-value compounds from brown seaweeds. In *Advances in Botanical Research*: *Vol. 95*. *Seaweeds around the world: state of art and perspectives* (pp. 369–413). Elsevier. https://doi.org/10.1016/bs.abr.2019.11.011

Contreras, L., Ritter, A., Dennett, G., Boehmwald, F., Guitton, N., Pineau, C., … Correa, J. A. (2008). Two-dimensional gel electrophoresis analysis of brown algal protein extracts. *Journal of Phycology*, *44*(5), 1315–1321. https://doi.org/10.1111/j.1529-8817.2008.00575.x

de Moura, J. M. L. N., Campbell, K., Mahfuz, A., Jung, S., Glatz, C. E., and Johnson, L. (2008). Enzyme-Assisted Aqueous Extraction of Oil and Protein from Soybeans and Cream De-emulsification. *Journal of the American Oil Chemists' Society*, *85*(10), 985–995. https://doi.org/10.1007/s11746-008-1282-2

Dobrinčić, A., Balbino, S., Zorić, Z., Pedisić, S., Bursać Kovačević, D., Elez Garofulić, I., and Dragović-Uzelac, V. (2020). Advanced technologies for the extraction of marine brown algal polysaccharides. *Marine Drugs*, *18*(3). https://doi.org/10.3390/md18030168

Duarte, C. M., Bruhn, A., and Krause-Jensen, D. (2021). A seaweed aquaculture imperative to meet global sustainability targets. *Nature Sustainability*, *5*(3), 185–193. https://doi.org/10.1038/s41893-021-00773-9

DuBois, Michel., Gilles, K. A., Hamilton, J. K., Rebers, P. A., and Smith, Fred. (1956). Colorimetric Method for Determination of Sugars and Related Substances. *Analytical Chemistry*, *28*(3), 350–356. https://doi.org/10.1021/ac60111a017

El-Aneed, A., Cohen, A., and Banoub, J. (2009). Mass spectrometry, review of the basics: electrospray, MALDI, and commonly used mass analyzers. *Applied Spectroscopy Reviews*, *44*(3), 210–230. https://doi.org/10.1080/05704920902717872

FAO of the UN. (2021). *WAPI factsheet to facilitate evidence-based policy- making and sector management in aquaculture*.

Filote, C., Santos, S. C. R., Popa, V. I., Botelho, C. M. S., and Volf, I. (2021). Biorefinery of marine macroalgae into high-tech bioproducts: a review. *Environmental Chemistry Letters*, *19*(2), 969–1000. https://doi.org/10.1007/s10311-020-01124-4

Fleurence, J. (1999). The enzymatic degradation of algal cell walls: a useful approach for improving protein accessibility? *Springer Science and Business Media LLC*. https://doi.org/10.1023/a:1008183704389

Flórez-Fernández, N., Torres, M. D., González-Muñoz, M. J., and Domínguez, H. (2019). Recovery of bioactive and gelling extracts from edible brown seaweed Laminaria ochroleuca by non-isothermal autohydrolysis. *Food Chemistry*, *277*, 353–361. https://doi.org/10.1016/j.foodchem.2018.10.096

Gaiero, M., Pham, T. T., Barile, D., and de Moura Bell, J. M. L. N. (2023). Kelping it Green: Innovative Analytical Tools for Eco-Friendly Giant Kelp Processing. U.S. Department of Energy AlgaePrize Competition.

Garcia-Vaquero, M., Rajauria, G., O'Doherty, J. V., and Sweeney, T. (2017). Polysaccharides from macroalgae: Recent advances, innovative technologies and challenges in extraction and purification. *Food Research International (Ottawa, Ont.)*, *99*(Pt 3), 1011–1020. https://doi.org/10.1016/j.foodres.2016.11.016

Global Market Insights. (2021, July 28). Commercial Seaweed Market size to exceed $95bn by 2027. Retrieved March 14, 2023, from https://www.gminsights.com/pressrelease/commercial-seaweed-market

Gordalina, M., Pinheiro, H. M., Mateus, M., da Fonseca, M. M. R., and Cesário, M. T. (2021). Macroalgae as protein sources—a review on protein bioactivity, extraction, purification and characterization. *Applied Sciences*, *11*(17), 7969. https://doi.org/10.3390/app11177969

Hahn, T., Lang, S., Ulber, R., and Muffler, K. (2012). Novel procedures for the extraction of fucoidan from brown algae. *Process Biochemistry*, *47*(12), 1691–1698. https://doi.org/10.1016/j.procbio.2012.06.016

Haroun-Bouhedja, F., Ellouali, M., Sinquin, C., and Boisson-Vidal, C. (2000). Relationship between sulfate groups and biological activities of fucans. *Thrombosis Research*, *100*(5), 453–459. https://doi.org/10.1016/s0049-3848(00)00338-8

Huang, Y.-P., Robinson, R. C., and Barile, D. (2022). Food glycomics: Dealing with unexpected degradation of oligosaccharides during sample preparation and analysis. *Journal of Food and Drug Analysis*, *30*(1), 62–76. https://doi.org/10.38212/2224-6614.3393

Huang, Y.-P., Robinson, R. C., Dias, F. F. G., de Moura Bell, J. M. L. N., and Barile, D. (2022). Solid-Phase Extraction Approaches for Improving Oligosaccharide and Small Peptide Identification with Liquid Chromatography-High-Resolution Mass Spectrometry: A Case Study on Proteolyzed Almond Extract. *Foods*, *11*(3), 340. https://doi.org/10.3390/foods11030340

IMARC group. (2021). Seaweed Market Size, Share, Value, Price Trends, Industry Analysis & Forecast (2023-2028). Retrieved March 13, 2023, from https://www.imarcgroup.com/seaweed-market

Jia, R.-B., Li, Z.-R., Lin, L., Luo, D., Chen, C., and Zhao, M. (2022). The potential mechanisms of Macrocystis pyrifera polysaccharides mitigating type 2 diabetes in rats. *Food & Function*, *13*(14), 7918–7929. https://doi.org/10.1039/d2fo01083f

Jordan, P., Fromme, P., Witt, H. T., Klukas, O., Saenger, W., and Krauss, N. (2001). Three-dimensional structure of cyanobacterial photosystem I at 2.5 A resolution. *Nature*, *411*(6840), 909–917. https://doi.org/10.1038/35082000

Kadam, S. U., Tiwari, B. K., and O'Donnell, C. P. (2015). Extraction, structure and biofunctional activities of laminarin from brown algae. *International Journal of Food Science & Technology*, *50*(1), 24–31. https://doi.org/10.1111/ijfs.12692

Koh, H. S. A., Lu, J., and Zhou, W. (2019). Structure characterization and antioxidant activity of fucoidan isolated from Undaria pinnatifida grown in New Zealand. *Carbohydrate Polymers*, *212*, 178–185. https://doi.org/10.1016/j.carbpol.2019.02.040

Kostas, E. T., Adams, J. M. M., Ruiz, H. A., Durán-Jiménez, G., and Lye, G. J. (2021). Macroalgal biorefinery concepts for the circular bioeconomy: A review on biotechnological developments and future perspectives. *Renewable and Sustainable Energy Reviews*, *151*, 111553. https://doi.org/10.1016/j.rser.2021.111553

Lee, Y.-K., Lim, D.-J., Lee, Y.-H., and Park, Y.-I. (2006). Variation in Fucoidan Contents and Monosaccharide Compositions of Korean Undaria pinnatifida (Harvey) Suringar (Phaeophyta). *ALGAE*, *21*(1), 157–160. https://doi.org/10.4490/ALGAE.2006.21.1.157

Lin, J. D., Lemay, M. A., and Parfrey, L. W. (2018). Diverse Bacteria Utilize Alginate Within the Microbiome of the Giant Kelp Macrocystis pyrifera. *Frontiers in Microbiology*, *9*, 1914. https://doi.org/10.3389/fmicb.2018.01914

Lloyds Register Foundation. (n.d.). *Seaweed Revolution: a manifesto for a sustainable future*. Retrieved from https://www.seaweedmanifesto.com/

Marinho, G. S., Holdt, S. L., and Angelidaki, I. (2015). Seasonal variations in the amino acid profile and protein nutritional value of Saccharina latissima cultivated in a commercial IMTA system. *Journal of Applied Phycology*, *27*(5), 1991–2000. https://doi.org/10.1007/s10811-015-0546-0

Mayo Clinic. (2023, July 1). Acarbose (Oral Route) Proper Use. Retrieved August 1, 2023, from https://www.mayoclinic.org/drugs-supplements/acarbose-oral-route/proper-use/drg-20067949

Michel, G., Tonon, T., Scornet, D., Cock, J. M., and Kloareg, B. (2010). The cell wall polysaccharide metabolism of the brown alga Ectocarpus siliculosus. Insights into the evolution of extracellular matrix polysaccharides in Eukaryotes. *The New Phytologist*, *188*(1), 82–97. https://doi.org/10.1111/j.1469-8137.2010.03374.x

Nagai, K., Yotsukura, N., Ikegami, H., Kimura, H., and Morimoto, K. (2008). Protein extraction for 2-DE from the lamina of Ecklonia kurome (laminariales): recalcitrant tissue containing high levels of viscous polysaccharides. *Electrophoresis*, *29*(3), 672–681. https://doi.org/10.1002/elps.200700461

Nagao, R., Yokono, M., Teshigahara, A., Akimoto, S., and Tomo, T. (2014). Light-harvesting ability of the fucoxanthin chlorophyll a/c-binding protein associated with photosystem II from the Diatom Chaetoceros gracilis as revealed by picosecond time-resolved fluorescence spectroscopy. *The Journal of Physical Chemistry. B*, *118*(19), 5093–5100. https://doi.org/10.1021/jp502035y

Pereira, L. (2021). Macroalgae. *Encyclopedia*, *1*(1), 177–188. https://doi.org/10.3390/encyclopedia1010017

Pérez, M. J., Falqué, E., and Domínguez, H. (2016). Antimicrobial Action of Compounds from Marine Seaweed. *Marine Drugs*, *14*(3). https://doi.org/10.3390/md14030052

Pharr, D. M., and Dickinson, D. B. (1973). β-Glucosidase contaminant in enzymatic reagent for glucose detection. *Analytical Biochemistry*, *51*(1), 315–319. https://doi.org/10.1016/0003-2697(73)90480-6

PhycoTerra. (2022, September 12). Unique Benefits of Macroalgae & Microalgae in Agriculture. Retrieved July 18, 2023, from https://phycoterra.com/blog/benefits-of-macroalgae-and-microalgae-in-agriculture/

Rioux, L. E., Turgeon, S. L., and Beaulieu, M. (2007). Characterization of polysaccharides

extracted from brown seaweeds. *Carbohydrate Polymers*, *69*(3), 530–537. https://doi.org/10.1016/j.carbpol.2007.01.009

Saravana, P. S., Cho, Y.-N., Woo, H.-C., and Chun, B.-S. (2018). Green and efficient extraction of polysaccharides from brown seaweed by adding deep eutectic solvent in subcritical water hydrolysis. *Journal of Cleaner Production*, *198*, 1474–1484. https://doi.org/10.1016/j.jclepro.2018.07.151

Sari, Y. W., Mulder, W. J., Sanders, J. P. M., and Bruins, M. E. (2015). Towards plant protein refinery: Review on protein extraction using alkali and potential enzymatic assistance. *Biotechnology Journal*, *10*(8), 1138–1157. https://doi.org/10.1002/biot.201400569

Shi, X., Huang, Y., Mao, Y., Naimy, H., and Zaia, J. (2012). Tandem mass spectrometry of heparan sulfate negative ions: sulfate loss patterns and chemical modification methods for improvement of product ion profiles. *Journal of the American Society for Mass Spectrometry*, *23*(9), 1498–1511. https://doi.org/10.1007/s13361-012-0429-4

Souza Almeida, F., Furlan Goncalves Dias, F., Kawazoe Sato, A. C., and Leite Nobrega de Moura Bell, J. M. (2021). From solvent extraction to the concurrent extraction of lipids and proteins from green coffee: An eco-friendly approach to improve process feasibility. *Food and Bioproducts Processing*, *129*, 144–156. https://doi.org/10.1016/j.fbp.2021.08.004

Vásquez, V., Martínez, R., and Bernal, C. (2019). Enzyme-assisted extraction of proteins from the seaweeds Macrocystis pyrifera and Chondracanthus chamissoi: characterization of the extracts and their bioactive potential. *Journal of Applied Phycology*, *31*(3), 1–12. https://doi.org/10.1007/s10811-018-1712-y

Vishchuk, O. S., Ermakova, S. P., and Zvyagintseva, T. N. (2011). Sulfated polysaccharides from brown seaweeds Saccharina japonica and Undaria pinnatifida: isolation, structural characteristics, and antitumor activity. *Carbohydrate Research*, *346*(17), 2769–2776. https://doi.org/10.1016/j.carres.2011.09.034

Wang, S.-H., Huang, C.-Y., Chen, C.-Y., Chang, C.-C., Huang, C.-Y., Dong, C.-D., and Chang, J.-S. (2020). Structure and Biological Activity Analysis of Fucoidan Isolated from Sargassum siliquosum. *ACS Omega*, *5*(50), 32447–32455. https://doi.org/10.1021/acsomega.0c04591

Wijesinghe, W. A. J. P., and Jeon, Y.-J. (2012). Enzyme-assistant extraction (EAE) of bioactive components: a useful approach for recovery of industrially important metabolites from seaweeds: a review. *Fitoterapia*, *83*(1), 6–12. https://doi.org/10.1016/j.fitote.2011.10.016

Wong, P.-F., Tan, L.-J., Nawi, H., and AbuBakar, S. (2006). Proteomics of the red alga, gracilaria changii (gracilariales, rhodophyta). *Journal of Phycology*, *42*(1), 113–120. https://doi.org/10.1111/j.1529-8817.2006.00182.x

Yong, W. T. L., Thien, V. Y., Rupert, R., and Rodrigues, K. F. (2022). Seaweed: A potential climate change solution. *Renewable and Sustainable Energy Reviews*, *159*, 112222. https://doi.org/10.1016/j.rser.2022.112222

Yuan, Y., Zhang, J., Fan, J., Clark, J., Shen, P., Li, Y., and Zhang, C. (2018). Microwave assisted extraction of phenolic compounds from four economic brown macroalgae species

and evaluation of their antioxidant activities and inhibitory effects on α-amylase, α-glucosidase, pancreatic lipase and tyrosinase. *Food Research International (Ottawa, Ont.)*, *113*, 288–297. https://doi.org/10.1016/j.foodres.2018.07.021

Zhang, Z., Khan, N. M., Nunez, K. M., Chess, E. K., and Szabo, C. M. (2012). Complete monosaccharide analysis by high-performance anion-exchange chromatography with pulsed amperometric detection. *Analytical Chemistry*, *84*(9), 4104–4110. https://doi.org/10.1021/ac300176z

Zhan, Y., Marchand, C. H., Maes, A., Mauries, A., Sun, Y., Dhaliwal, J. S., … Zerges, W. (2018). Pyrenoid functions revealed by proteomics in Chlamydomonas reinhardtii. *Plos One*, *13*(2), e0185039. https://doi.org/10.1371/journal.pone.0185039