Title

Advancing Analytical Methodologies to Study Oligosaccharides, Bioactive peptides, and Allergenic Proteins in Plant-based Beverages and Related Products

Author

Huang, Yu-Ping

Publication Date

2022

Peer reviewed|Thesis/dissertation

Advancing Analytical Methodologies to Study Oligosaccharides, Bioactive peptides, and Allergenic Proteins in Plant-based Beverages and Related Products

By

YU-PING HUANG
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Food Science

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Daniela Barile, Chair

David A. Mills

Juliana Maria Leite Nobrega de Moura Bell

Committee in Charge

2022

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to the people who have helped me during the journey toward finishing this dissertation.

First, I gratefully thank Dr. Daniela Barile, my doctoral advisor, for her guidance and tremendous support over the past five years. I am grateful for her kindness, patience, and continuous encouragement, which strongly supported my research, career, and life. Her insightful thoughts and enthusiasm have inspired me in many ways and broadened my vision of science.

I want to thank my dissertation committee members, Dr. David Mills and Dr. Juliana Maria Leite Nobrega de Moura Bell, for their advice and encouragement. The valuable feedback they provided greatly improved this dissertation.

I would like to thank all the Barile Lab current and previous members for their help with my research and life. My special appreciation goes to Dr. Tian Tian, Dr. Apichaya (Garn) Bunyatratchata, and Dr. Randall Robinson for mentoring me when I started working in the Barile Lab. They always generously shared their experience and helped me solve problems. Their help enabled me to initiate my dissertation studies smoothly.

I want to express my appreciation to all the collaborators who have contributed to the studies in this dissertation. A special thank goes to Dr. Fernanda Furlan Goncalves Dias for her support with several relevant projects. Thanks to Dr. Naomi K. Fukagawa, Dr. Katherine M. Phillips, Kazunori Machida, Dr. Judy Van de Water, and Joseph Schauer for their contribution. The collaborations allowed me to expand the horizon of this dissertation.

Lastly, I would like to thank my husband, Dr. Austin Horng-En Wang, who has given me the most vital support.

ABSTRACT

As the global consumption of plant-based beverages and related food products made from various crops has been rapidly growing, the contribution of plants as a source of various nutrients needs to be assessed and compared with traditional sources such as dairy milk. Proteins and carbohydrates are two classes of macronutrients; besides providing energy and building blocks for biosynthesis, they may exert additional biological activities in the human body. Several peptides derived from food digestion (or food processing) have been reported to exert specific bioactivities, thereby beneficially affecting human health. Indigestible carbohydrates, especially oligosaccharides that are low in molecular weight, hold the potential to reshape the gut microbiome via prebiotic activity and lead to various health-promoting effects. The bioactivities of peptides and oligosaccharides are governed by their structures, such as the amino acid sequences, constituent monomers, glycosidic linkages, and molecular sizes. On the other hand, some plant proteins are known food allergens; yet, their allergenicity might be reduced during food processing by altering allergenic proteins' structures. Characterizing the structures of these molecules at different processing stages can improve our understanding of the potential bioactivities of specific foods and guide the development of optimized processing conditions. Analytical approaches based on liquid chromatography and mass spectrometry are useful to achieve in-depth characterization of all these compounds. This dissertation presents seven case studies about the analysis of bioactive molecules and protein allergens in plant-based products using glycomics, peptidomics, and proteomics techniques.

Chapter I reviews the current knowledge on the nutritional and bioactive properties of protein and carbohydrate components in plant-based beverages. Chapters II to IV focus on the development and optimization of analytical methods for performing high-quality glycomics and

peptidomics in complex food matrices, and provides solutions to overcome current challenges. Chapter IV presents a workflow for discovering small bioactive peptides in plant-based foods by LC-MS/MS. The innovative application of the dimethyl labeling technique in food peptidomics effectively facilitated full-length sequencing of small peptides with two to four amino acid residues that are believed to exert more potent *in vivo* bioactivities than larger peptides.

Chapters V to VII presents further applications of the optimized glycomics and peptidomics analytical methods on plant-based beverages and related products generated by various processing approaches. Chapter V describes a comprehensive analysis of oligosaccharides in almond milk, soy milk, and soy flour. Chapter VI demonstrates the application of the optimized glycomics method to identify naturally occurring oligosaccharides in chickpeas as well as new structures generated from polysaccharides breakdown operated by enzymes. Chapter VII describes the discovery of potentially bioactive oligosaccharides and peptides in the cooking water of chickpeas and common beans, which is known as aquafaba, with our optimized glycomics and peptidomics approaches. It was found for the first time that when peptides were dimethyl labeled, α- and γ-glutamyl peptides could be easily differentiated with the uniquely significant a_1 and b_1 fragment ions. Based on that, many γ-glutamyl peptides with potential kokumi and anti-inflammatory activities were identified in aquafaba.

Chapter VIII investigates the effect of enzymatic treatments for decreasing almond protein allergenicity, using proteomics analysis and immunoassay. Enzymatic extraction using neutral protease significantly reduced immunoglobulin E- and immunoglobulin G-reactivities, as evidenced by immunoblotting using human sera from patients allergic to almonds. The results were supported by proteomics, which revealed that a majority of almond proteins were hydrolyzed by neutral protease during the enzymatic extraction; however, the β-subunit regions in prunin 1

and prunin 2, which are constituents of a major almond allergen—amandin, showed resistance to proteolysis by neutral protease. Proteomics analysis also confirmed that the linear epitopes in the β-subunit regions in prunin 1 and prunin 2 largely kept their integrity.

This work provides innovative and optimized analytical approaches for characterizing food-derived oligosaccharides and bioactive peptides as well as demonstrating practical applications of the optimized methodologies to various plant-based foods. The in-depth characterization of oligosaccharides, peptides, and allergenic proteins offers key insights into strategies to optimize processing conditions, valorize low-cost streams, and enhance the nutritional values of food products.

RELEVANT PUBLICATIONS

Huang, Y.-P., Robinson, R. C., Dias, F. F. G., de Moura Bell, J. M. L. N., & 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

Huang, Y.-P., Dias, F. F. G., de Moura Bell, J. M. L. N., & Barile, D. (2022). A complete workflow for discovering small bioactive peptides in foods by LC-MS/MS: A case study on almonds. *Food Chemistry*, *369*, 130834. https://doi.org/10.1016/j.foodchem.2021.130834

Huang, Y.-P., Robinson, R. C., & 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

Machida, K., **Huang, Y.-P.**, Furlan Gonçalves Dias, F., Barile, D., & 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

Huang, Y.-P., Paviani, B., Fukagawa, N.K., Phillips, K.M., & Barile, D. Comprehensive oligosaccharide profiling of commercial almond milk, soy milk, and soy flour. *Food Chemistry* (under review).

Dias, F. F. G., **Huang, Y.-P.**, Schauer J., Barile, D., Van de Water, J., & de Moura Bell, J. M. L. N. Effects of proteolysis on almond protein profile, digestibility, and antigenicity (manuscript in preparation).

Huang, Y.-P. & Barile, D. Don't throw away the cooking water: Aquafaba from chickpeas and common beans contains potentially bioactive oligosaccharides and peptides (manuscript in preparation).

Huang, Y.-P., de Moura Bell, J. M. L., & Barile, D. Proteins and carbohydrates in plant-based beverages: Chemical composition, nutritional values, and potential bioactivities (manuscript in preparation).

TABLE OF CONTENTS

Chapter I

Proteins and carbohydrates in plant-based beverages: Chemical composition, nutritional values, and potential bioactivities

Abstract

Global consumption of plant-based beverages has been growing rapidly in recent years. As a result, novel foods have been developed to meet consumers' demands. In many western countries, consumers drink plant-based beverages as an alternative to cow's milk due to various considerations. Although food manufacturers have been successful at mimicking the appearance and texture of milk, the chemical composition and nutritional values remain substantially different. This chapter presents an overview of the compositional and nutritional properties of proteins and carbohydrates in some mainstream and emerging plant-based beverages, including soy, almond, rice, oat, and pea milk. The composition of various plant-based beverages can vary considerably, but overall, they have much lower protein content than cow's milk, except for soy milk and some pea milk products. Unlike cow's milk and other animal proteins, which provide sufficient essential amino acids that meet humans' needs, most plant-based beverages lack one or more essential amino acids. A notable exception is soy milk, which is considered a source of complete protein for consumers above three years of age. The potential issue of consuming incomplete protein arises primarily because the labels of novel products tend to promote the high protein content instead of focusing on protein quality, and so fail to mention the absence of essential amino acids that are required for the correct functioning of the body. Individuals at a particularly high risk of consuming products lacking essential amino acids include growing children, the elderly, and pregnant women, all of whom have an increased protein requirement. Some proteinaceous compounds found in the raw materials used for producing plant-based beverages, such as lectins, Bowman-Birk inhibitors, lunasin, and some low-molecular-weight peptides, possess specific bioactivities beneficial or detrimental to human health; the current understanding of the characteristics of these compounds was summarized and discussed. Unsweetened plant-based beverages made from soybeans,

almonds, and peas contain lower levels of available carbohydrates than cow's milk, whereas cereal-based beverages usually contain higher levels of digestible carbohydrates due to the high starch content in cereal grains. Varying levels of dietary fiber with diverse structures are present in plant-based beverages. Some of the classes of fiber possess potential bioactivities, such as the prebiotic activities of raffinose family oligosaccharides and the cholesterol-lowering property of β-glucans. The available knowledge of proteins, peptides, and carbohydrates in plant-based beverages is primarily extrapolated from the corresponding raw material crops. However, processing can considerably affect their chemical composition and molecular structures and, consequently, their bioactive functions. Characterizing the many bioactive molecules in commercial products and emerging plant-based beverages being produced with new materials or innovative processing methods will help understand their potential health benefits for consumers and thus guide the selection of raw materials and processing strategies to produce foods with desired nutritional and functional properties.

1.1. Introduction

Plant-based beverages are made from various raw material crops, including legumes, cereals, tree nuts, nuts, and others. In several East Asian countries, where the lactase-persistence trait is at low frequency (Lomer, Parkes, & Sanderson, 2008), plant-based beverages such as soy milk are traditional drinks that have long been widely consumed. In contrast, in North America and Europe, plant-based beverages have just started to become popular in recent years and are usually considered analogs—in terms of appearance and texture—as well as dietary substitutes for cow's milk. Many factors may influence consumers' choice between cow's milk and plant-based beverages, such as dietary restrictions, health considerations, sustainability, culture, animal welfare, and affordability. For example, plant-based beverages are the preferred choice for consumers suffering from lactose intolerance or allergies to cow's milk, as they are free of lactose and milk proteins that could induce allergic reactions in sensitive individuals. In recent years, consumers have become more aware of the potential health benefits of plant-based diets and the importance of reducing the environmental impact caused by animal food production. Despite the significant controversy among experts regarding the metrics to quantify greenhouse gases (GHG), and the difficulty to obtain reliable measures to achieve an equitable comparison of the impacts of innovative and traditional protein products (Liu, Proudman, & Mitloehner, 2021), the newly developed awareness about reducing the impact on the environment seems to be one of the incentives for consumers to increase the consumption of plant-based beverages. Younger adults generally are paying more attention to the health of the planet and the impact of food production on the environment (Pew Research Center, 2021). A recent report by Innova Market Insights revealed that consumers want to be both ethically and environmentally conscious, and so when purchasing food and beverages, they are ranking the planet's health higher than their own health

(Innova Market Insights, 2021). Therefore, many consumers who do not suffer from milk allergies/intolerance and are not vegan are still choosing to reduce their consumption of animal protein in favor of plant protein.

However, it is important to know that while food manufacturers tend to formulate their new products to match the consistency and taste of milk, the composition of plant-based beverages and cow's milk varies considerably due to their distinct origins. Therefore, although plant-based beverages are presented to consumers as dietary alternatives to cow's milk, they are not interchangeable with respect to their nutritional properties. Even within plant-based beverages, the variety of raw material crops and manufacturing procedures results in rather varied compositions and nutritional properties. Generally, plant-based beverages are produced through a series of unit operations, including soaking, wet milling, filtration, formulation with added ingredients, homogenization, and heat treatment pasteurization or sterilization (Aydar, Tutuncu, & Ozcelik, 2020; Mäkinen, Wanhalinna, Zannini, & Arendt, 2016). The soaking and the subsequent wet milling steps might be replaced by dry milling followed by extraction (Mäkinen et al., 2016), depending on the design of the manufacturing process. Similarly, the processing steps' sequence, approaches, and settings (e.g., temperature, time, and pressure) often differ among manufacturers, creating variation among products. The manufacturing process of plant-based beverages involves particle size reduction through milling and extraction of various components with water from the plant materials. The efficacy of milling and extraction can significantly affect the compositional and nutritional properties of the products and likely can have an impact on their bioactivities. The subsequent separation of the extract from the insoluble fraction also eliminates part of the nutrients and makes the product composition differentially deviate from the raw material crops. Nonetheless, whole-grain beverages are sometimes produced to keep all the constituents of the

whole seeds by skipping filtration, along with proper processing techniques, such as high-pressure processing and media-milling (Kuo, Chen, & Yeh, 2014; Li et al., 2021). The starting materials may also vary, considering the many available plant varieties, international sourcing, and growing conditions, consequently affecting the composition of the final beverages. For example, rice milk can be made from either brown rice or milled rice (white rice). During rice milling, germ and bran that are enriched in oil, protein, dietary fiber, vitamins, minerals, and other phytochemicals are removed from brown rice (Eyarkai Nambi, Manickavasagan, & Shahir, 2017; Huang & Lai, 2016). Therefore, rice milk made from brown rice and rice bran would generally contain higher amounts of the components in rice bran mentioned above than the beverage made from white rice. Pea milk is an emerging plant-based beverage typically made from pea protein isolate or concentrate, currently being manufactured and available on the market in a few regions, including Europe, North America, and Australia. The use of pea protein isolate or concentrate effectively increased the protein content of pea milk, while the composition of pea protein isolate or concentrate, and consequently the pea milk, could be significantly affected by the pea cultivars and extraction methods for making pea protein (Cui et al., 2020; Stone, Avarmenko, Warkentin, & Nickerson, 2015; Yang, Zamani, Liang, & Chen, 2021).

The consumption of plant-based milk has been steadily increasing and has gradually become an essential part of the diets of particular groups of people. Plant-based beverage drinkers among Canadians increased from 1.8% to 3.0% from 2004 to 2015, with a considerable decline in plain dairy milk consumers during the same period (Islam, Shafiee, & Vatanparast, 2021). A study based on a United States household survey in 2019 showed that 22.8% of households consumed almost exclusively plant-based beverages, and 15.6% of households frequently consumed both dairy milk and plant-based beverages (Wolf, Malone, & McFadden, 2020). The global market of

plant-based beverages is projected to proliferate at a compound annual growth rate of 14.3% from 2021 to 2028 (Grand View Research, n.d.).

Due to the increased consumption of plant-based beverages and the massive compositional difference between cow's milk and various plant-based beverage products, it is crucial to scrutinize their compositions to better understand their influence on human nutrition and health. Either too little or too much protein both have devastating consequences on health, and evidence of broader damage resulting from excessive consumption of protein can be seen by the planet's languishing resources.

This chapter reviews the protein and carbohydrate components in the mainstream plant-based beverages available on the market, juxtaposing them to enable comparisons and a thorough understanding of their nutritional and bio-functional properties.

1.2. Proximate composition

Table 1.1 shows the proximate composition of commercial cow's milk and plant-based beverages in the United States. The fat content in cow's milk is typically standardized during processing and classified as whole-fat, reduced-fat (2% fat), low-fat (1%), and nonfat milk. The overall composition of milk may also slightly vary with the fat content. For example, the average moisture content of milk in the USDA food composition database (Foundation Foods) ranges from 88.1 (in whole milk) to 90.8% (in nonfat milk) in milk with a fat content varying from high to low, and an average protein content varying from 3.3 to 3.4% (FoodData Central, 2022). The carbohydrate content in milk is around 5%, which is mainly attributed to lactose. The average lactose content ranges from 4.8 (in whole milk) to 5.1% (in nonfat milk) (FoodData Central, 2022).

Table 1.1. Proximate composition of plant-based beverages and cow's milk (g/100 g).[1]

Beverage	Moisture	Protein	Fat	Ash	Carbohydrate (by difference)
Almond milk					
unsweetened, plain, shelf-stable	97.4 (97.1–97.6)	0.55 (0.44–0.69)	1.22 (1.03–1.6)	0.49 (0.22–0.7)	0.34
unsweetened, plain, refrigerated	96.5 (93.7–97.5)	0.66 (0.44–1.5)	1.56 (1.15–1.98)	0.6 (0.21–0.93)	0.67
Soy milk					
unsweetened, plain, shelf stable	92.4 (90.3–93.6)	3.55 (3–4.69)	2.12 (1.86–2.7)	0.64 (0.46–0.81)	1.29
unsweetened, plain, refrigerated	91.5 (90.4–91.8)	2.78 (2.56–3.38)	1.96 (1.4–2.88)	0.75 (0.7–0.89)	3
Rice milk[2]					
unsweetened	89.3	0.28 (0–0.42)	0.97 (0.83–1.04)	0.3	9.17 (8.33–9.58)
Oat milk					
unsweetened, plain, refrigerated	90.6 (87.8–95.1)	0.8 (0.38–1.19)	2.75 (0.22–4.68)	0.79 (0.44–0.96)	5.1
Pea milk[3]					
unsweetened, original		3.33	1.88		0
2% reduced fat		1.67	2.08		1.67
Cow's milk					
whole, 3.25% milkfat	88.1 (87.4–89.2)	3.27 (2.93–3.51)	3.2 (2.59–3.41)	0.8 (0.6–1.25)	4.63
reduced fat, 2% milkfat	89.1 (88–90.7)	3.36 (3–3.7)	1.9 (1.68–2.06)	0.75 (0.48–1.1)	4.9
lowfat, 1% milkfat	89.7 (86.9–91.8)	3.38 (3.06–4.02)	0.95 (0.72–1.19)	0.8 (0.52–1.55)	5.18
Nonfat (skim)	90.8 (89.3–92.8)	3.43 (3–4.15)	0.08 (0.03–0.16)	0.77 (0.5–1.3)	4.92

[1] Data source: (FoodData Central, 2022).

[2] Data based on label claim or estimated.

[3] Data based on label claims of two individual branded samples.

The chemical composition of plant-based beverages can diverge considerably due to various factors, including the variation of plant materials, the processing techniques, and the formulation. Among a few mainstream plant-based beverages (almond, soy, rice, and oat milk), the average moisture content in products sold in the United States ranges from 89.3 to 97.4%, with

almond milk having the highest value (i.e., most dilute) and rice milk having the lowest, according to the available data in the USDA food composition database (FoodData Central, 2022). In other words, the average dry matter content ranges from 2.6 to 10.7%, representing a difference of up to four times.

The composition of the dry matter also differs across various plant-based beverages. In general, soy milk has the highest protein content (2.56–4.69%), with an average of 3.55% and 2.78% for shelf-stable and refrigerated products, respectively. Almond milk, rice milk, and oat milk contain less protein (0.44–1.5%, 0–0.42%, and 0.38–1.19%, respectively) than soy milk (Table 1.1 and Fig. 1.1). According to the nutritional labels, the protein content in two commercial pea milk products was 1.67 and 3.33%. The substantial difference between them reflects that the materials, processing, and formulation can vary considerably, even for a plant-based beverage made from the same crop. Among different plant-based beverages, only soy milk, and potentially pea milk, offer a similar protein content to cow's milk.

Carbohydrate content, which is usually calculated by difference based on other proximate compositions, also varies among various plant-based beverages. Rice milk displayed the highest average carbohydrate content (9.17%), followed by oat milk (5.1%) (Table 1.1 and Fig. 1.1). In comparison, soy milk and almond milk contain lower carbohydrates (0.46–0.89% and 0.21–0.93%, respectively). The lipid and ash contents range from 0.22–2.88% and 0.21–0.96%, respectively, both of which do not considerably deviate from the contents in cow's milk and are often adjusted due to externally added vegetable oil and fortified minerals.

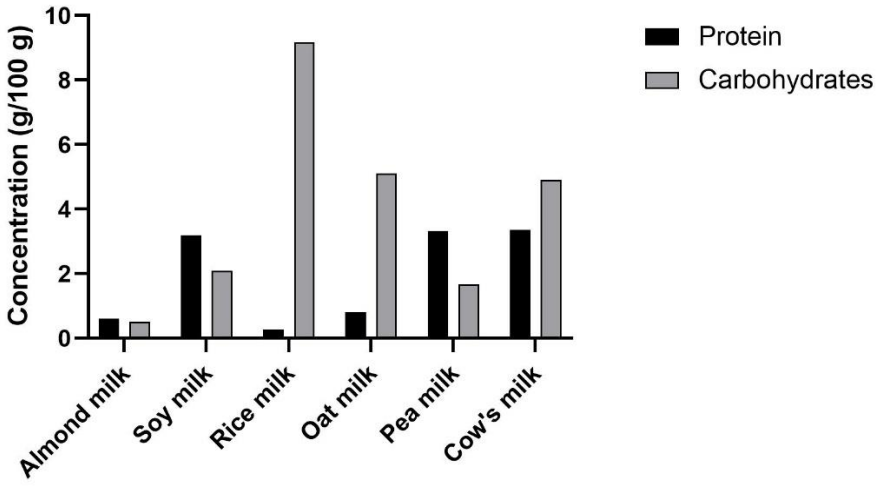


Fig. 1.1. Protein and carbohydrate contents in commercial unsweetened plant-based beverages and cow's milk in the United States.

Note: Data source: (FoodData Central, 2022). Almond milk and soy milk include both shelf-stable and refrigerated products. Cow's milk includes whole, reduced fat, low fat, and nonfat milk. Sample size: almond milk, n = 16; soy milk, n = 15, rice milk, n = 3; oat milk, n = 8; pea milk, n = 1, cow's milk, n = 96. Because the data of some individual samples are not available, the standard deviations was not calculated.

Except for the properties of plant materials (e.g., composition and extractable components), the proximate composition of plant-based beverages could be greatly affected by the manufacturing process, such as the ratio of plant materials to water, the extraction efficiency, and upstream treatments to which the plant materials might have been subjected (e.g., heat treatments). The price of plant materials can be a major factor influencing the beverage formulation. For example, almond seeds are more pricey than other raw materials such as rice and soybeans. The price is likely one of the driving factors limiting the relative amount of almond seeds being used

in almond milk formulation. Dry matter content that can be dissolved or stably suspended in water, possibly assisted by adding emulsifiers and gums, may also restrict the amount of plant materials being added. Plant-based beverages sold in Europe provide information about the ratio of plant materials used in the ingredients, which could render some hints about the commercial products' formulation. According to a major almond milk manufacturer's website, almond materials make up only 2% of the ingredients in their unsweetened almond milk product (containing 0.45% protein and 0.2% carbohydrates) (Almond Breeze, 2022). The website of a company in the UK selling various plant-based beverages also reveals that significantly different percentages of plant materials are used for producing various beverages (e.g., 12.5% rice for rice milk, 9.8% oats for oat milk, 8.7% soybeans for soy milk, and 2.3% almonds for almond milk) (Alpro, 2022). Based on the information above, it can be inferred that the low protein content in commercial almond milk is mainly caused by the low amounts of almonds being added as an ingredient (e.g., almond paste) rather than a true issue of protein extractability from the source. Given that almond milk is the most desirable plant-based beverage for consumers and the most widely consumed in the United States among the various plant-based beverages offerings (McCarthy, Parker, Ameerally, Drake, & Drake, 2017; Wolf et al., 2020), consumers need to be aware of the low protein content to prevent insufficient nutrient intake when almond milk is used in full substitution of cow's milk.

Although the proximate composition sketches the fundamental nutritional values of plant-based milk, because the composition of protein, carbohydrates, and fat are essentially different between cow's milk and plant-based beverages and even among different plant-based beverages, it is important to understand the compositional properties to delineate the roles of these nutrients for human health. The following sections examine the chemical composition and nutritional properties of proteins and carbohydrates in plant-based beverages and the related bioactivities.

1.3. Protein

1.3.1. Amino acid composition

Amino acids can be acquired from dietary proteins following food digestion and absorption. They can provide energy, serve as materials for protein synthesis, and play various roles in maintaining human body functions. Traditionally, amino acids are categorized into nutritionally essential (indispensable) and nonessential (dispensable) based on whether the amino acids' carbon skeleton can be synthesized endogenously in the body (Wu, 2010). Essential amino acids (histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, and valine) cannot be synthesized by the human body and thus must be obtained from diets. Table 1.2 shows the amino acid profiles of various plant materials used for making plant-based beverages. Compared with cow's milk, the plant materials contain lower percentages of essential amino acids (27.58–39.58% vs. 41.4%). Among the five plant-based materials, soybeans (39.57%) and pea protein isolate (39.58%) contain the highest percentages of essential amino acids, followed by rice (37.1%), pea protein concentrate (36.21%), oats (36.2%), and almonds (27.58%). Among the essential amino acids, branched-chain amino acids (BCAA), which include leucine, isoleucine, and valine, can exert several important functions in the human body, such as stimulating protein synthesis, inhibiting proteolysis, and modulating neurotransmission (Holeček, 2018). Most food materials used for making plant-based beverages, including soybeans, rice, oats, and pea protein isolate, have a percentage of BCAA to total amino acids (17.92, 18.4, 17.4, and 16.65%, respectively) that are comparable to cow's milk (17.7%). Almonds are the only exception and contain only 12.75% of BCAA. Methionine and cysteine are the only two amino acids containing sulfur on the side chain among the 20 canonical amino acids present in proteins. These sulfur amino acids (SAA) play unique roles in human body functions. For example, methionine is the

amino acid initiating protein translation; cysteine can be used for synthesizing glutathione, a peptide defending against oxidative stress, and stabilizing protein structures by forming disulfide bonds with other cysteine residues (Brosnan & Brosnan, 2006). Cereal is usually rich in SAA and often used to complement the low SAA content in legumes for vegan eaters. Oats and rice contain the highest percentage of SAA to total amino acids (4.4 and 3.4%, respectively) among the foods used for making plant-based beverages (1.5–4.4%) and cow's milk (2.4%), agreeing with the general tendency of cereals. Soybeans (although belonging to legumes) and almonds have a similar or slightly higher percentage of SAA (2.78 and 2.35%, respectively) compared with cow's milk. The percentages of SAA to total amino acids in pea protein concentrate (1.2%) and pea protein isolate (1.5%) are significantly lower than the other food materials.

Table 1.2. Amino acid composition of proteins in various foods.[1]

Amino acid	Soybean	Almond	Brown rice	Oat	Pea protein		Cow's milk
					Conc.	Isolate	
Alanine	4.41 (4.34–4.48)	4.28 (4.16-4.39)	5.84	4.9	4.13	4.41	3.0
Arginine	7.45 (7.28–7.72)	10.53 (10.48–10.62)	7.58	7.3	14.22	7.15	3.3
Aspartic acid	11.37 (11.27–11.46)	11.26 (10.91–11.71)	9.37	8.5	11.58	11.52	7.8
Cystine	1.37 (1.32–1.49)	1.34 (1.21–1.43)	1.22	2.7	0.35	0.73	0.6
Glutamic acid	18.31 (18.11–18.39)	26.73 (26.27–27.13)	20.3	20.7	16.39	17.03	23.1
Glycine	4.32 (4.27–4.36)	6.88 (6.73–7.09)	4.93	5.0	4.50	4.68	1.8
Histidine	2.96 (2.90–2.98)	2.04 (1.99–2.12)	2.55	2.6	3.39	3.81	3.0
Isoleucine	4.90 (4.84–4.99)	2.92 (2.74–3.09)	4.23	4.0	3.48	3.68	4.2
Leucine	7.96 (7.86–8.03)	6.50 (6.42–6.57)	8.28	7.8	6.94	8.16	8.7
Lysine	6.63 (6.51–6.73)	2.96 (2.82–3.05)	3.82	4.4	8.12	8.96	8.1
Methionine	1.41 (1.37–1.44)	1.01 (0.95–1.08)	2.26	1.7	0.85	0.78	1.8
Phenylalanine	5.29 (5.19–5.38)	5.04 (4.96–5.14)	5.16	5.5	4.67	5.18	4.8
Proline	4.96 (4.89–5.05)	4.15 (4.11–4.27)	4.69	5.5	4.30	5.01	9.6
Serine	4.60 (4.52–4.68)	4.41 (4.29–4.60)	5.18	5.3	4.96	6.09	4.8
Threonine	3.87 (3.82–3.97)	2.85 (2.82–2.90)	3.67	3.6	3.12	3.69	4.5
Tryptophan	1.50 (1.42–1.55)	0.93 (0.87–1.01)	1.27	0.9	0.51	0.51	1.5
Tyrosine	3.65 (3.52–3.77)	2.85 (2.65–2.98)	3.75	3.8	3.35	3.79	4.5
Valine	5.07 (5.04–5.12)	3.32 (3.15–3.42)	5.87	5.5	5.13	4.81	4.8
Total EAA%	39.57 (39.37–39.76)	27.58 (27.32–27.99)	37.1	36.2	36.21	39.58	41.4
Total BCAA%	17.92 (17.80–18.14)	12.75 (12.44–12.98)	18.4	17.4	15.55	16.65	17.7
Total SAA%	2.78 (2.69–2.93)	2.35 (2.17–2.49)	3.47	4.4	1.2	1.5	2.4
Reference	(Lagos & Stein, 2017)	(House, Hill, Neufeld, Franczyk, & Nosworthy, 2019)	(FoodData Central, 2022)	(Pettersson, Lindberg, Thomke, & Eggum, 1996)	(Pettersson et al., 1996)	(Tömösközi, Lásztity, Haraszi, & Baticz, 2001)	(Guo et al., 2007)

[1] Expressed as % of total amino acids.

1.3.2. Protein quality

Protein quality is usually defined by amino acid profiles and protein digestibility, which inform about the amount and the composition of amino acids that foods can provide to the human body. Protein digestibility-corrected amino acid score (PDCAAS) has been used in the United States for evaluating dietary protein quality for human nutrition and was adopted by the Food and Agriculture Organization of the United Nations (FAO) in 1989 (Food and Agriculture Organization of the United Nations/World Health Organization, 2013; Marinangeli & House, 2017). More recently, the digestible indispensable amino acid score (DIAAS) was recommended by an FAO Expert Consultation in 2011 as a new protein quality scoring system, attempting to address the limitations of PDCAAS (Food and Agriculture Organization of the United Nations/World Health Organization, 2013). DIAAS uses true ileal digestibility (TID) of individual amino acids, instead of fecal digestibility of crude protein used by PDCAAS, to better approach the actual amino acid utilization. Table 1.3 summarizes the protein digestibility and DIAAS of food materials commonly used for making plant-based beverages. In general, the protein digestibility of legumes (soybean and pea, total amino acid TID = 92–99% and 92–98%, respectively) is in the similar range to cow's milk (94–99%) but higher than oats (84–87%), rice (77–82%), and almonds (true fecal protein digestibility = 86–90%). Although legumes often contain protease inhibitors, it appears that the effect of protease inhibition on digestibility is near negligible, possibly because protease inhibitors in the tested legume foods might be well inactivated or removed by processing. Despite the similar digestibility of some plant foods with cow's milk, the DIAAS of all the plant foods in Table 1.3 are all lower than cow's milk.

DIAAS are usually determined for three different age groups (<0.5 yr, 0.5–3 yr, and >3 yr) according to the specific amino acid requirements. Infants (<0.5 yr), followed by young children

(0.5–3 yr), require higher levels of proteins and amino acids to support their growth aside from the maintenance than older children, adolescents, and adults, whose protein and amino acid requirements are primarily used for maintenance (Food and Agriculture Organization of the United Nations/World Health Organization, 2013; Millward, 1997). Accordingly, the DIAAS of a given food is always the lowest in the <0.5 yr age group and the highest in the >3 yr age group. Typically, infants from birth to six months are exclusively fed with breast milk or infant formula, so they are not consumers of plant-based beverages. Therefore, the discussion here focuses on the 0.5–3 yr and >3 yr age groups. The DIAAS of cow's milk for the 0.5–3 yr and >3 yr age groups are both well above 100, indicating that proteins in cow's milk meet all the essential amino acid requirements of the age groups and even can balance other dietary proteins that are deficient in any essential amino acids. Among all the plant materials, soybeans have the highest DIAAS with the value of 84–99 and 98–117 for the 0.5–3 yr and >3 yr age groups, respectively. Thus, for the >3 yr age group, soybeans are a source of complete protein and, similar to cow's milk, can potentially complement other incomplete dietary proteins. In contrast, the other plant materials generally have inferior DIAAS, within the ranges of ~40–60 for the 0.5–3 yr age group and ~45–70 for the >3 yr group. According to the FAO expert consultation report (Food and Agriculture Organization of the United Nations/World Health Organization, 2013), it is recommended that DIAAS cut-off values of 100 and 75 could be used for determining protein quality categories of excellent and good, respectively. By following this rule, soybeans may be claimed as good and excellent protein sources for the 0.5–3 yr and >3 yr age groups, respectively, whereas other plant materials generally cannot be claimed as excellent nor good protein sources. Therefore, pea milk, oat milk, rice milk, and almond milk individually cannot be used as a dietary alternative to cow's

milk, nutritionally, when considering the protein quality, not to mention the low quantity of protein being supplied in some plant-based beverages, as discussed above.

A feasible approach to improve protein quality and raise DIAAS for plant-based beverages is to combine various protein sources with complementary digestible indispensable amino acid (DIAA) compositions. SAA and Lys are frequently determined as the first limiting amino acid in legumes and cereals, respectively, in line with their deficiency in the particular crops. It is a well-known concept that legumes (rich in lysine but lacking SAA) and cereals (rich in SAA but lacking lysine) can complement the amino acid compositions of each other when they are consumed together. Calculating DIAAS for beverage products made with multiple crops will inform the effectiveness of combining various protein sources for protein quality enhancement. In general, protein foods from an individual source that have the lower DIAA reference ratio (compared with the amino acid requirement patterns) for their limiting amino acids will need other foods with the higher DIAA reference ratio with appropriate quantity to reach the ideal DIAAS.

Table 1.3. Protein digestibility and DIAAS of food materials for making plant-based beverages and cow's milk.

Plant source	Food type	Digestibility%[a]		Animal model[b]	Conversion factor[c]	DIAAS (first limiting amino acid)			References
		Range of individual AA	Total AA			<0.5 yr	0.5–3 yr	>3 yr	
Soybean	Soy milk	84.0 (Arg)–104.6 (Cys)	92.3	Adult Yucatan miniature pigs	6.25	78 (Leu)	99 (Lys)	117 (Val)	(Reynaud et al., 2021)
	Soy flour	81 (Cys)–99 (Arg)	95	Growing pigs (barrows)	6.25	73 (Leu)	89 (SAA)	105 (SAA)	(Mathai, Liu, & Stein, 2017)
	Soy protein isolate	91 (Cys)–101 (Arg)	99	Growing pigs (barrows)	6.25	68 (SAA)	84 (SAA)	98 (SAA)	(Mathai et al., 2017)
	Soy protein isolate	90 (Trp)–98 (Glu, Tyr, His, Lys)	97	Growing male rats	6.25	73.5 (SAA)	89.8 (SAA)	105 (SAA)	(Rutherfurd, Fanning, Miller, & Moughan, 2015)
	Soy protein isolate	89 (Met)–98 (Glu, Lys, Arg)	96	Growing male rats	6.25	74.1 (SAA)	90.6 (SAA)	106 (SAA)	(Rutherfurd et al., 2015)
Pea	Pea emulsion[d]	80.3 (Met)–102.7 (Pro)	94.2	Adult Yucatan miniature pigs	6.25	42 (SAA)	51 (SAA)	60 (SAA)	(Reynaud et al., 2021)
	Pea protein concentrate	75 (Cys)–99 (Arg)	96	Growing pigs (barrows)	6.25	45 (Trp)	62 (SAA)	73 (SAA)	(Mathai et al., 2017)
	Pea protein concentrate	91 (Trp)–98 (Glu, Ala, Cys, Ile, Leu, Try, Phe, Lys, Arg)	98	Growing male rats	6.25	67.3 (SAA)[e]	82.2 (SAA)	96.5 (SAA)[e]	(Rutherfurd et al., 2015)
	Cooked pea	83 (Trp)–96 (Lys)	92	Growing male rats	6.25	47.4 (SAA)[e]	57.9 (SAA)	68.0 (SAA)[e]	(Rutherfurd et al., 2015)
Oat	Cooked rolled oats	79 (Trp)–93 (Cys)	87	Growing male rats	6.25	44.8 (Lys)[e]	54.2 (Lys)	64.4 (Lys)[e]	(Rutherfurd et al., 2015)
	Oat protein concentrate	72.0 (Asp) –95.0 (Trp)	84.2	Growing pigs (barrows)	Not specified	41 (Phe+Tyr)	56 (Lys)	67 (Lys)	(Abelilla, Liu, & Stein, 2018)
Rice	Cooked rice	55 (Met) –92 (Lys)	77	Growing male rats	6.25	49.2 (Lys)[e]	59.5 (Lys)	70.7 (Lys)[e]	(Rutherfurd et al., 2015)

	Rice protein concentrate	63 (Met) –91 (Arg)	82	Growing male rats	6.25	30.6 (Lys)[e]	37.1 (Lys)	44.1 (Lys)[e]	(Rutherfurd et al., 2015)
Almond	Ground almond	-	85.7–89.9[f]	Rats	5.18	38.6 (Lys)[g]	46.7 (Lys)[g]	55.4 (Lys)[g]	(House et al., 2019)
					6.25	32.0 (Lys)[g]	38.7 (Lys)[g]	45.9 (Lys)[g]	(House et al., 2019)
Cow's milk	Skimmed milk powder	73 (Cys)–98 (Arg)	94	Growing pigs (barrows)	6.25	81 (Thr)	105 (SAA)	123 (SAA)	(Mathai et al., 2017)
	Milk protein concentrate	85 (Cys)–117 (Gly)	99	Growing pigs (barrows)	6.25	85 (Trp)	120 (SAA)	141 (SAA)	(Mathai et al., 2017)
	Milk protein concentrate	92 (Met, Ile, Trp)–99 (Tyr, Phe, Lys)	94	Growing male rats	6.25	96.5 (SAA)[e]	118 (SAA)	139 (SAA)[e]	(Rutherfurd et al., 2015)

[a] Digestibility refers to true ileal digestibility (TID) unless otherwise specified. Individual AA: individual AA digestibility. Total AA: total amino acid digestibility.

[b] Animal model for digestibility measurement.

[c] Conversion factor for estimating crude protein content.

[d] Pea emulsion refers to a soybean oil-in-water emulsion containing a pea isolate.

[e] The DIAAS values were obtained by conversion based on the literature data (Rutherfurd et al., 2015) and FAO-recommended amino scoring patterns (Food and Agriculture Organization of the United Nations/World Health Organization, 2013).

[f] The value represents the true fecal protein digestibility of four almond varieties.

[g] The DIAAS values were calculated according to FAO (Food and Agriculture Organization of the United Nations/World Health Organization, 2013) based on the amino acid composition, protein content, and protein digestibility data in the literature (House et al., 2019) and FAO recommended amino scoring patterns. The value represents the average of four almond varieties.

It is noteworthy that a recent review article by Craddock, Genoni, Strutt, & Goldman (2021) raised some concerns about using DIAAS to evaluate the protein quality of plant-based foods. A primary concern that may compromise the correctness of comparison across different foods is using the generalized nitrogen-to-protein conversion factor (6.25) in protein content determination. The food-specific conversion factors of plant-based foods are generally lower than in cow's milk due to the differences in amino acid compositions (Jones, 1931; Mariotti, Tomé, & Mirand, 2008). The relatively low specific conversion factors may lead to the overestimation of protein content in plant foods when a generalized conversion factor is used. However, they may also cause an underestimation of DIAAS and DIAA reference ratios. For example, the specific conversion factor for almonds was determined as 5.18, based on the nitrogen content (19.3%) in an almond protein amandin. The DIAAS and DIAA reference ratios calculated using the specific conversion factor will be 17.1% higher than the values obtained using the generalized factor of 6.25. The calculation of DIAA in almonds, based on the data provided by House et al. (2019), using 5.18 and 6.25 for nitrogen-to-protein conversion, resulted in DIAAS 45.9 (with five limiting amino acids) and 55.4 (with seven limiting amino acids), respectively, for the >3 yr age group.

FAO also recommended considering the availability of lysine while measuring DIAAS (Food and Agriculture Organization of the United Nations/World Health Organization, 2013). The high reactivity of lysine with reducing sugars in the Maillard reaction may cause the loss of available lysine during thermal processing and storage. Cow's milk is enriched with lactose (near 5%, w/w) (FoodData Central, 2022), which is a good reactant in the Maillard reaction, especially during thermal processing. In comparison, most sugars in soybean seeds are non-reducing sugars (e.g., sucrose (~4.7%) and stachyose (~3.2%)) (Hou et al., 2009). Reducing sugars that can participate in the Maillard reaction are low in abundance in soybeans (~0.5% glucose and ~0.4%

fructose) (A. Hou et al., 2009). The reduction of lysine in percentages caused by Maillard reaction in unsweetened soy milk is expected to be less significant than in cow's milk, although it still depends on the plant varieties and sources, specific processing procedures, and storage conditions.

1.3.3. Protein composition

The materials for making plant-based beverages are generally seeds and, therefore, rich in seed storage proteins. Seed storage proteins are accumulated as protein bodies in specific organelles in mature seeds and supply amino acids during seed germination and seedling growth (Shewry, Napier, & Tatham, 1995). Due to the high abundance of seed storage proteins, their composition is prominently responsible for the proteins' properties in physical functionality and the values in human nutrition. Table 1.4 displays the major seed storage proteins in crops used for making plant-based beverages. Although the crop plants are distantly related taxonomically, the first or the second most abundant storage proteins in the crops reviewed here all belong to the 11–12S globulin family (also called legumin-type globulins), which has a non-covalently bonded hexameric structure with a molecular weight of ~300 to 380 kDa. These include glycinin in soybean, amandin in almonds, glutelin in rice, and 12S globulin or avenalin in oat (as the most abundant storage protein), and legumin in pea (as the second most, sometimes the most, abundant storage protein) (Gueguen & Barbot, 1988; Lam, Can Karaca, Tyler, & Nickerson, 2016). Each subunit of the hexamers comprises an acidic (α; ~32–46 kDa) and a basic (β; ~17–22 kDa) subunits linked via an interchain disulfide bond (Albillos et al., 2008; D. H. Hou & Chang, 2004; Muench & Okita, 1997; Singh Sindhu, Zheng, & Murai, 1997; Wen & Luthe, 1985). The acidic and basic subunits were post-translationally released from the N- and C-terminuses, respectively, of the precursor proteins (Krishnan, 2000). Although most cereals accumulate prolamin as the primary storage protein (e.g., wheat, barley, and maize), rice and oats are two exceptions that contain

mainly 11–12S globulins, which account for 60–65% and 50–80%, respectively, of the total protein (Kawakatsu & Takaiwa, 2019; Klose & Arendt, 2012; Shewry & Halford, 2002).

Table 1.4. Protein composition of various raw material crops for making plant-based beverages and cow's milk.

Food	Major protein	MW (kDa)	Abundance	Reference
Soy	11S (glycinin)	300–380	52%	(Fujiwara, Hirai, Chino, Komeda, & Naito, 1992; Kinsella, 1979; Krishnan, 2000; Maruyama et al., 1998)
	Acidic	37–44		
	Basic	17–22		
	7S (β-conglycinin)	150–200	35%	
	α'	76		
	α	72		
	β	53		
	15S		5%	
	2S		5%	
Rice	Glutelin	~350	60–65%	(Kawakatsu, Hirose, Yasuda, & Takaiwa, 2010; Kawakatsu & Takaiwa, 2019; Wakasa, Yang, Hirose, & Takaiwa, 2009)
	Acidic	35		
	Basic	22		
	Prolamin	13 24	~20%	
	α-Globulin		5–10%	
Almond	Amandin/prunin/almond major protein	370	~70%	(Albillos et al., 2008; Sathe et al., 2002)
	Prunin-1	61		
	Prunin-2	55.9		
	Acidic	42–46		
	Basic	20–22		
Pea	Legumin (11S)	320–380	33.4–41.8%	(Barac et al., 2010; Lam et al., 2016)
	Acidic	40		
	Basic	20		
	Vicilin (7S)	150–170	44.3–49.9%	
	α, β, and γ	12–36		
	Convicilin	~70	9.5–12.3%	
	Albumin	5–80	10–20%	
Oat	12S globulin (avenalin)	322	50–80%	

	α	31.7		(Klose &
	β	21.7		Arendt, 2012;
	7S			Peterson,
	3S			1978)
	Prolamin		4–15%	
	Albumin		1–12%	
	Glutelin		<10%	
Cow's milk	α_{S1}-Casein	23.6	32%	(Walstra,
	α_{S2}-Casein	25.2	8.4%	Walstra,
	β-Casein	24.0	26%	Wouters, &
	κ-Casein	19.6	9.3%	Geurts, 2005)
	γ-Casein	20.5	2.4%	
	β-Lactoglobulin	18.3	9.8%	
	α-Lactalbumin	14.2	3.7%	

7S globulin is the other storage protein rich in soybeans (β-conglycinin, 150–200 kDa) and peas (vicilin, 150–170 kDa), accounting for 35% and 44–50% (usually the most abundant in pea), respectively, of the total protein. Soybean β-conglycinin and pea vicilin are both trimeric proteins consisting of subunits of 76 (α'), 72 (α), and 53 (β) kDa and ~50 kDa, respectively (Gatehouse, Croy, Morton, Tyler, & Boulter, 1981; Maruyama et al., 1998). Smaller subunits of 12.5–35 kDa can further be formed from the ~50 kDa subunits that have been assembled into pea vicilin (Gatehouse et al., 1981); the two proteins are both glycoproteins with various levels of N-glycosylation (Gatehouse et al., 1981; Maruyama et al., 1998). In contrast to 11S globulins, 7S globulins contain low levels of SAA (i.e., methionine and cysteine), which may reduce the protein quality (Krishnan, 2000; Lam et al., 2016). The generally higher relative abundance of 7S globulins than 11S globulins in peas than soybeans is also associated with the lower percentage of SAA in peas.

1.3.4. Allergenic proteins

Consumption of certain foods may trigger allergic responses in sensitive individuals. Milk, soybeans, and tree nuts, along with eggs, fish, shellfish, peanuts, wheat, and sesame, belong to the

nice major food allergens designated by the Food Allergen Labeling and Consumer Protection Act (FALCPA) and Food Allergy Safety, Treatment, Education, and Research (FASTER) Act of the United States (Benedict, 2006). Therefore, besides cow's milk, some plant-based beverages, including soy milk and almond milk, can potentially cause food allergies. According to a few population-based surveys of adults in North America, the prevalence of self-reported food allergy to tree nuts (0.9–1.3%) and soybeans (0.1–0.6%) are slightly lower than milk (1.9–4.1%) (Messina & Venter, 2020), although the prevalence may vary when the population dietary habits alter.

The initiation of allergic reaction involves sensitization at first exposure to allergens, which leads to the generation of immunoglobulin E (IgE) by B cells, and manifestation at re-exposure, which is mediated by IgE and/or T cells and leads to allergic symptoms (Valenta, Hochwallner, Linhart, & Pahr, 2015). World Health Organization and International Union of Immunological Societies (WHO/IUIS) Allergen Nomenclature Sub-committee (http://allergen.org/) (Radauer et al., 2014) has identified several allergenic proteins in soybeans and almonds. The allergenic proteins include Gly m1 (hydrophobic protein from soybean), Gly m 3 (profilin), Gly m 4 (pathogenesis-related protein, PR-10), Gly m 5 (β-conglycinin), Gly m 6 (glycinin), Gly m 7 (seed biotinylated protein), and Gly m 8 (2S albumin) in soybeans and Pru du 1 (PR-10), Pru du 3 (non-specific lipid transfer protein 1, nsLTP1), Pru du 4 (profilin), Pru du 5 (60s acidic ribosomal prot. P2), Pru du 6 (amandin), Pru du 8 (antimicrobial seed storage protein), and Pru du 10 (mandelonitrile lyase 2) in almonds. Even though peas are not a major food allergen, three allergenic proteins, namely Pis s 1 (vicilin), Pis s 2 (convicilin), and Pis s 3 (nsLTP), were still recognized by WHO/IUIS. Even though some potential allergenic proteins or peptides in rice and oats were reported (Real et al., 2012; Trcka et al., 2012), they were not documented as food allergens in the WHO/IUIS database. Several highly abundant storage proteins in crops used for

making plant-based beverages are allergenic proteins, such as amandin in almonds and glycinin and β-conglycinin in soybeans; it was found that IgE reactivity to these proteins was a potential indicator of almond allergy and severe soy allergy, respectively (Holzhauser et al., 2009; Ito et al., 2011; Kabasser et al., 2021).

Epitopes, which play essential roles in allergenic reactions, are specific regions of allergenic proteins that can be recognized by antibodies or antigen receptors (Liu & Sathe, 2018). They can be linear sequences composed of several amino acids (linear or sequential epitopes) or nonconsecutive amino acids or peptide sequences that are in close proximity (conformational or nonsequential epitopes) (Liu & Sathe, 2018). Allergenicity of protein allergens could be altered by food processing when epitope structures are modified. Dhakal et al. (2014) reported that high-pressure processing significantly reduced the immunoreactivity of almond milk to linear and conformational epitopes on amandin, whereas thermal processing was only effective when reaching specific temperatures and an extended holding time. Thermal processing, fermentation, and enzymatic processing were shown to be effective in reducing soy protein allergenicity (Pi, Sun, Fu, Wu, & Cheng, 2021). It is often difficult to precisely predict the effect of food processing on allergenic protein structures. Although food allergenicity is often studied with immuno-based assays, such as immunoblot and enzyme-linked immunosorbent assay (ELISA), incorporating mass spectrometry-based proteomics approaches in the relevant studies can allow further understanding of structural and sequential alterations of allergenic proteins.

1.3.5. Proteins with biological activities

Aside from supplying amino acids, some proteins (and peptides) are resistant to gastrointestinal digestion. These proteins may exert certain biological activities, which could be beneficial or detrimental to the human body in their functionally active form.

1.3.5.1. Lectins

Lectins are a group of proteins that can bind to specific mono- or oligosaccharides reversibly and are often found in significant abundance in legumes, including the materials for making plant-based beverages, soybeans (accounting for ~1–2% of seed protein) and peas (de Mejia, Bradford, & Hasler, 2003; Lajolo & Genovese, 2002; Vasconcelos & Oliveira, 2004). Lectins from various plants have different binding specificities. Soybean lectin binds to N-acetyl-D-galactosamine (D-GalNAc) and D-galactose, whereas pea lectin binds to α-D-glucose and α-D-mannose (Lis & Sharon, 1973). N-acetyl-D-glucosamine (D-GlcNAc)-specific lectins are present in several kinds of cereals, including rice and other plants, but absent in oats (Mishkind, Palevitz, Raikhel, & Keegstra, 1983; Vasconcelos & Oliveira, 2004). Lectins isolated from rice bran exhibited a stronger lectin-carbohydrates interaction with the oligomers of D-GlcNAc (degree of polymerization, DP = 3 to 6) than the dimer (di-N-acetylchitobiose) followed by the monomer (Nakata et al., 2017).

The carbohydrate-binding property of lectins also enables them to bind to specific glycans on the surface of small intestinal enterocytes, leading to subsequent endocytosis into the epithelial cells (Pusztai et al., 1993). Adverse effects related to plant lectin consumption observed in animals include growth inhibition (Jindal, Soni, & Singh, 1982; Pusztai et al., 1990), a decrease in digestive enzyme activity (Jindal et al., 1982), small intestine growth due to hyperplasia and hypertrophy (Lajolo & Genovese, 2002; Pusztai et al., 1993, 1990), and pancreas enlargement (Grant, Alonso, Edwards, & Murray, 2000; Grant, Dorward, & Pusztai, 1993; Jordinson et al., 1996). The severity of the resulting deleterious effects may vary depending on the carbohydrate-binding specificities of different lectins and their abundances. On the other hand, despite the known adverse effects, some beneficial properties of lectins were also demonstrated. For example, soybean and rice bran

lectins were shown to exhibit *in vitro* antitumor activity (de Mejia et al., 2003; Miyoshi et al., 2001) and modulate transepithelial transport across human intestinal Caco-2 cell monolayers (Yamamoto et al., 2013).

As mentioned, lectins are often resistant to proteolysis by human digestive enzymes, such as pepsin, trypsin, and chymotrypsin (Lajolo & Genovese, 2002; Muramoto, 2017; Poola, 1989). Nevertheless, lectins are heat-labile and thus can be inactivated via thermal processing under proper conditions. It was reported that aqueous heat treatment at 60 °C for 40 min and 75 °C for 2 h did not affect the activity of soybean and rice lectins, respectively, so harsher conditions were necessary for the inactivation (Armour, Perera, Buchan, & Grant, 1998; Poola, 1989). For example, aqueous heat treatment of soybean at 100 °C for 10 min led to a complete loss of lectin activity (Armour et al., 1998). Complete inactivation of lectins in vegetable pea seeds was achieved through ordinary cooking for 20 min, autoclaving at 121 °C for 10 min, or microwaving for 4 min (Habiba, 2002). The production of plant-based beverages usually includes thermal processing for pasteurization or sterilization, which could partially or fully inactivate lectins in the plant seeds. Detectable levels of active lectins were found in commercial soy milk sold in Italy and Mexico with concentrations of 4.7 μg/mL and 6.91–16.20 μg/g, respectively (de la Barca, Vázquez-Moreno, & Robles-Burgueño, 1991; Rizzi et al., 2003).

1.3.5.2. Protease inhibitors

Protease inhibitors are widely found in plant seeds, including legumes and cereals. Active protease inhibitors may impair protein digestion in the gastrointestinal tract by inhibiting the activity of digestion enzymes in humans, specifically trypsin, chymotrypsin, and elastin, and thus, are usually considered antinutritional factors. Among various legumes and cereal grains, soybean has the highest trypsin inhibitory activity (Rackis, Wolf, & Baker, 1986). Soybean protease

inhibitors are the most extensively studied to date. Two main types of protease inhibitors are present in soybean seeds, namely Bowman-Birk protease inhibitors (BBIs) and the Kunitz trypsin inhibitor. BBIs are also present in several legumes, including peas, common beans, chickpeas, lentils, and some cereals, such as wheat, rice, and barley. Generally, legumes contain substantially higher amounts of protease inhibitors than cereals (Rackis et al., 1986). Thus, cereals' protease inhibitors are of less concern regarding their antinutritional effect.

BBIs are small cysteine-rich proteins containing several disulfide bridges and one to two protease inhibitory sites in each ~8 kDa protein molecule or domain (Prakash et al., 1996). The structural stability of BBIs against proteolysis by digestive enzymes is associated with the high density of disulfide bridges. Soybean BBIs have a few isoforms with the molecular weights of ~8 kDa (71–76 amino acids), containing seven disulfide bridges and two specific binding sites for trypsin and chymotrypsin (IBB1, UniProt accession P01055), two trypsin molecules (IBBD2, UniProt accession P01064), or elastase and chymotrypsin (IBBC2, UniProt accession P01063) resulting in protease inhibition (Baek & Kim, 1993; Baek, Song, Choi, & Kim, 1994; Odani & Ikenaka, 1977). Like soybean BBIs, pea BBIs also include several isoforms with molecular weights of 7–8 kDa and seven intra-chain disulfide bonds (Domoney, Welham, Sidebottom, & Firmin, 1995; Ferrasson, Quillien, & Gueguen, 1995; Quillien, Ferrasson, Molle, & Gueguen, 1997). Two reactive sites involved in trypsin and chymotrypsin inhibition are present in pea BBIs (Clemente, Gee, Johnson, Mackenzie, & Domoney, 2005). Rice BBIs also exhibit inhibitory activities to trypsin and chymotrypsin (J. Chen et al., 2006). They exist in various forms consisting of one (8 kDa), two (16 kDa), or three domains (25 kDa); each domain contains four or five disulfide bridges (Chen et al., 2006; Lin et al., 2006). Protease inhibitors in rice seeds are mainly located in the embryo, whereas the endosperm has no detectable protease inhibitory activity

(Horiguchi & Kitagishi, 1971). Therefore, protease inhibitors may only exist in rice milk made of brown rice but not white rice (milled rice) that only consists of the endosperm.

Besides acting as an antinutritional factor, several beneficial properties of BBIs were also studied. Anticarcinogenic activity of soybean BBIs against various cancers, such as colorectal, prostate, and breast cancer, were demonstrated in *in vitro* and *in vivo* models (Gitlin-Domagalska, Maciejewska, & Dębowski, 2020). Like soybean BBIs, pea BBIs were shown to inhibit the proliferation of HT29 human colorectal cancer cells in a dose-dependent manner without affecting the growth of non-malignant CCD-18Co colon cells (Clemente, Carmen Marín-Manzano, Jiménez, Carmen Arques, & Domoney, 2012; Clemente et al., 2005; Clemente, Moreno, Marín-Manzano, Jiménez, & Domoney, 2010). Although the mechanism of BBIs' anticarcinogenic activity has yet to be fully understood, such activity is thought to be associated with the inhibition of trypsin- and/or chymotrypsin-like serine proteases (Clemente et al., 2012, 2010). Due to the involvement of several serine proteases (e.g., cathepsin G, neutrophil elastase, and mast cell chymase) in the human body's inflammatory responses and auto-inflammatory reactions, BBIs also exhibited anti-inflammatory and immunoregulatory activities and has the potential to be used as an oral anti-inflammatory drug (Sadeghalvad, Mohammadi-Motlagh, Karaji, & Mostafaie, 2019; Safavi & Rostami, 2012).

Kunitz trypsin inhibitor (KTI) is a 20 kDa protein containing 181 amino acid residues and two disulfide bridges. KTI can cause potent inhibition of trypsin, by blocking the active site via the formation of a stable inhibitor-enzyme complex, and weak inhibition of chymotrypsin (Blow, Janin, & Sweet, 1974); both the trypsin-inhibiting site and a second reactive site can bind chymotrypsin five orders of magnitudes less tightly than trypsin does (Bösterling & Quast, 1981). KTI contributed 22.1–79.8% of the total trypsin inhibitory activity in soybean (Kumar, Rani,

Mittal, & Shuaib, 2019). Compared to BBIs and lectins, there are much fewer reports about the beneficial bioactivity of KTI to date.

Protease inhibitors also can be inactivated by heat. The efficacy of the heat inactivation of protease inhibitors depends on the holding temperature and time. Yuan et al. (Yuan, Chang, Liu, & Xu, 2008) compared different heating methods on the efficacy of trypsin inhibitor inactivation. When soy milk was boiled or steam injected at 100 °C, the residual trypsin activity gradually decreased within a 20- or 30-min period. Boiling at 100 °C for 30 min resulted in a residual trypsin activity of 7.7–10.7%. Higher temperatures in the range of 121 to 154 °C in ultrahigh-temperature (UHT) processing facilitated faster inactivation of trypsin inhibitors in soy milk (Kwok, Qin, & Tsang, 1993). A residual of 10% trypsin inhibitory activity was achieved by holding at 143 °C for 56 s or 154 °C for 23 s (Kwok et al., 1993). The activity of rice protease inhibitors in a crude extract remained unchanged after heat treatment at 70 °C for 30 min, whereas heating at a higher temperature at 100 °C for 30 min eliminated more than 75% of protease activity (Horiguchi & Kitagishi, 1971). Sample types (e.g., brown rice grains and rice extract) may also affect the inactivation rate of rice trypsin inhibitors. Bradbury et al. (Bradbury, Hammer, & Sugani, 1992) compared the residual activity of trypsin inhibitors in brown rice and a crude aqueous rice extract during thermal treatment at 90 °C and pH 7. They found that trypsin inhibitors present in brown rice were fully inactivated within 10 min; however, less than 20% of trypsin inhibitors in the crude extract were inactivated after 30 min, suggesting that the processing procedure of rice milk (i.e., whether the cooking process takes place on brown rice grains) may affect the residual protease inhibitor activity. Protease inhibitor concentrations in commercial plant-based beverage products may vary considerably. Arques et al. (2014) quantified BBIs and KTI in six commercial soy milk products sold in Spain, with the measured concentrations being 0.60–9.07 mg/100 mL of BBIs and

1.82–5.50 mg/100 mL of KTI. The BBI concentrations in twelve commercial soy milk products in the United States reported in another study varied widely, ranging from not detected to 55.9 mg/100 mL (Hernández-Ledesma, Hsieh, & de Lumen, 2009a). Intriguingly, it was estimated that a glass of 200 mL soy milk provides the chymotrypsin inhibitor activity high enough to exert anticarcinogenic effects in humans based on the dose used in animal models (Arques, Pastoriza, Delgado-Andrade, Clemente, & Rufián-Henares, 2016).

1.3.5.3. Lunasin

Lunasin, a bioactive peptide discovered more recently than lectins and BBIs, is the small subunit of soybean 2S albumin, consisting of 43 amino acid residues with a size of ~5.1 kDa. The antimitotic activity of lunasin in mammalian cells was firstly revealed in *lunasin* (*GM2S-1*)-transfected cells by (Galvez & de Lumen, 1999). Due to the potential of being used as a chemopreventive agent, studies on lunasin have expanded quickly since then and demonstrated a variety of beneficial bioactivities of this peptide, including cancer prevention, antioxidation, anti-inflammation, anti-cholesterol, and immune system regulation (Hsieh, Martínez-Villaluenga, de Lumen, & Hernández-Ledesma, 2018). The amino acid sequence of lunasin (^{1}SKWQHQQDSC ^{11}RKQLQGVNLT ^{21}PCEKHIMEKI ^{31}QGRGDDDDDD ^{41}DDD) is composed of four regions with different functions, including a fragment with unknown functions f(1–22), a helical chromatin-binding region f(23–32), an RGD motif that is associated with cell adhesion and internalization f(33–35), and a highly negatively charged tail containing eight consecutive aspartic acid residues that can bind core histones f(36–43) (Hernández-Ledesma, Hsieh, & de Lumen, 2009b). The unique structure of the peptide is related to its antimitotic property. In cells undergoing transformation, the binding of lunasin to deacetylated histones inhibits acetylation of histones H3

and H4 via competition with acetyltransferase, subsequently inducing cell cycle arrest and apoptosis (Hernández-Ledesma et al., 2009b).

Although lunasin had been reported to be detected in oats (Nakurte et al., 2013) and other cereal grains (Jeong, Lam, & de Lumen, 2002; Jeong et al., 2007; Nakurte et al., 2012), the presence of the identical sequence of lunasin in cereals was questioned, because of the lack of DNA encoding the sequence, and appeared to be false identification due to the insufficient specificity of the analytical methods (Alaswad & Krishnan, 2016; Dinelli et al., 2014; Mitchell, Lovegrove, & Shewry, 2013). In that sense, soy milk is the only widely available plant-based beverage that contains lunasin. The concentration of lunasin in soy milk was reported in some studies, with the range of 10.7–18.9 mg/100 mL in twelve commercial soy milk products in the United States (Hernández-Ledesma et al., 2009a) and 1.78–9.18 mg/100 mL and 2.92–9.05 mg/100 ml in twelve regular and seven organic soy milk samples, respectively, from various origins (Cavazos, Morales, Dia, & De Mejia, 2012).

Bioavailability is critical in determining the true bioactivity of bioactive compounds in the human body. After the consumption of soy milk or other soybean products containing lunasin, digestive enzymes may hydrolyze proteinaceous lunasin in the gastrointestinal tract. For this reason, some studies evaluated the stability of lunasin using *in vitro* digestion and Caco-2 cell monolayer models. It was shown that intact lunasin could partially survive the simulated gastric and intestinal digestion (Hernández-Ledesma et al., 2009a; Park, Jeong, & Lumen, 2007) as well as the incubation in the apical side of Caco-2 cell monolayers, which have peptidases on the brush border membrane (Fernández-Tomé, Sanchón, Recio, & Hernández-Ledesma, 2018). Studies also revealed that BBI and KTI play a critical role in protecting lunasin from being destroyed by digestive enzymes (Cruz-Huerta et al., 2015; Hernández-Ledesma et al., 2009a; Hsieh, Hernández-

Ledesma, Jeong, Park, & de Lumen, 2010; Park et al., 2007). Other studies demonstrated the absorption of lunasin into the circulation system after oral intake, evidenced by the detection of lunasin in the blood and various organs of rodent models (Hsieh et al., 2010) and the plasma of humans (Dia, Torres, De Lumen, Erdman, & De Mejia, 2009).

1.3.5.4. Low-molecular-weight bioactive peptides

Low-molecular-weight peptides generally refer to peptides composed of 2 to 20 amino acid residues, although the size range is not strictly defined. Depending on the structural features and physicochemical properties, some low-molecular-weight peptides exhibit specific bioactivities, such as anti-hypertensive, anti-diabetic, anti-cancer, antimicrobial, and immunomodulatory activities (Daliri, Oh, & Lee, 2017; Minkiewicz, Iwaniak, & Darewicz, 2019; Sánchez & Vázquez, 2017) and are termed bioactive peptides. Proteolysis is effective in generating low-molecular-weight peptides with potential bioactivities and may occur naturally in raw food materials and during food processing (e.g., fermentation and enzymatic treatments) and gastrointestinal digestion. Enzymatic treatment using proteolytic enzymes has been used in food processing for various purposes, including increasing protein extraction rate, enhancing protein digestibility, and reducing allergenicity (Bahna, 2008; Eriksen, 1983; Koopman et al., 2009; Souza, Dias, Koblitz, & M. L. N. de M. Bell, 2019). It was demonstrated that using proteases in soy milk production from soy flour greatly improved the protein and solid yields (Eriksen, 1983). Generating potentially bioactive peptides could be an additional benefit of using proteolytic enzymes in food processing.

Bioactive peptides derived from cow's milk has been broadly studied, encompassing various activities. For example, two bioactive peptides IPP and VPP, with their amino acid sequences encrypted in β-casein (IPP and VPP) and κ-casein (IPP only), were initially identified

in Japanese sour milk fermented with a starter culture containing *Lactobacillus helveticus* and *Saccharomyces cerevisiae* (Nakamura, Yamamoto, Sakai, Okubo, et al., 1995; Nakamura, Yamamoto, Sakai, & Takano, 1995). Their antihypertensive effect, mainly associated with the inhibition of angiotensin I-converting enzyme (ACE), has been investigated by *in vitro*, *in vivo*, and human studies (Li et al., 2019). Based on meta-analyses of randomized clinical trials, IPP and VPP lactotripeptides supplementation could significantly reduce blood pressure (Chanson-Rolle, Aubin, Braesco, Hamasaki, & Kitakaze, 2015; Fekete, Givens, & Lovegrove, 2015).

Previous studies identified many bioactive peptides derived from the plant materials used for making plant-based beverages and directly from the beverages. Several peptides derived from the major storage proteins of soybeans (i.e., glycinin and β-conglycinin) were reported to exhibit various bioactivities, including cholesterol- and triglyceride-lowering, ACE inhibition, and anti-diabetic (Chatterjee, Gleddie, & Xiao, 2018). Some peptides generated from soy milk by enzymatic hydrolysis using a protease (PROTIN SD-NY10) and fermentation using lactic acid bacteria were identified as ACE inhibitory peptides, including two newly identified sequences with IC_{50} (half maximal inhibitory concentration) <10 μM (FFYY (1.9 μM) and WHP (4.8 μM)) in the protease-processed soy milk (Tomatsu, Shimakage, Shinbo, Yamada, & Takahashi, 2013; Undhad Trupti, Das, Solanki, Kinariwala, & Hati, 2021). Hydrolysis of pea protein using the protease Alcalase also led to the formation of three multifunctional peptides (IR, KF, and EF) that exhibit the inhibitory activities against ACE, renin, and calmodulin-dependent phosphodiesterase 1 (Li & Aluko, 2010). Capriotti et al. (2015) conducted simulated digestion on soy protein and soy milk and characterized the peptides in the digesta. They found three peptides with known ACE inhibitory and/or antioxidative activities being released during the digestion, with many other longer peptides encrypting amino acid sequences with those activities in the digesta. The blood

pressure-lowering effect of the oral intake of protein hydrolysates derived from soybean, pea, or rice using different proteases was demonstrated in multiple animal experiments (Daliri et al., 2019; Li, Qu, Wan, & You, 2007; Li et al., 2011; Wu & Ding, 2001; Yang, Yang, Chen, & Chen, 2008) or human clinical studies (Kwak et al., 2013; Li et al., 2011).

Notably, some low-molecular-weight peptides are not generated from protein hydrolysis but are biosynthesized as free peptides in plants. γ-Glutamyl peptides found in several legume seeds were identified as "kokumi" substances, which can elicit a complex sensation of thickness, continuity, and mouthfuls when combined with other basic taste compounds (Li, Zhang, & Lametsch, 2020). It was reported that raffinose and stachyose (major oligosaccharides in soybeans) synergistically enhanced the kokumi sensation of soybean γ-glutamyl peptides, γ-EF and γ-EY (Shibata et al., 2017). The activation of calcium-sensing receptors (CaSR), which are found on the cell surface of various tissues, including taste buds and intestinal epithelial cells, was suggested to be related to γ-glutamyl peptides' kokumi characteristics and anti-inflammatory activity (Amino et al., 2016; Guha & Majumder, 2022; Juan Yang, Bai, Zeng, & Cui, 2019).

It is worthy to note that most bioactive peptides were initially identified via *in vitro* experiments and may not necessarily exhibit the same biological activity *in vivo* (Foltz, van der Pijl, & Duchateau, 2010; Sato, 2018). Bioactive peptides may exert their activity in the gastrointestinal tract or other organs after they enter the systemic circulation. Depending on the target organs, the bioavailability of peptides involves different factors. For peptides to function in the gastrointestinal tract, they must be released and/or survive during digestion. In comparison, for peptides exerting systemic activity, besides the factor of digestion, the efficiency of intestinal transepithelial transport and the structural stability in first-pass metabolism are also key parameters that affect bioavailability. The release of peptides is associated with proteolytic enzymes' substrate

specificity, such as pepsin cleaves at the carboxyl end of Leu and Phe, and trypsin cleaves at the carboxyl end of Lys and Arg (Udenigwe, Abioye, Okagu, & Obeme-Nmom, 2021). Interestingly, although proteolysis takes place in the gastrointestinal tract with a variety of proteases and peptidases, protein hydrolysis using exogenous proteases (e.g., Thermolysin and Alcalase) before the oral intake of pea or rice protein was demonstrated to be essential for the blood pressure-lowering effect in spontaneously hypertensive rats (Li et al., 2007; Li et al., 2011). This indicates that pre-hydrolysis using specific proteases may alter the bioavailability of anti-hypertensive peptides. The small intestine can absorb peptides via intracellular and extracellular routes. Due to the unique high absorption efficiency of peptides containing two or three amino acid residues through a peptide transporter PEPT1 on the intestinal brush border membrane, di-/tripeptides potentially have higher bioavailability than longer peptides (Shen & Matsui, 2017). The stability of bioactive peptides during digestion, absorption, and circulation is dependent on peptide sequences (Sato, 2018; Shen & Matsui, 2017). Although it was argued that the concentrations of specific individual bioactive peptides in plasma were often at low levels after the oral intake (Foltz et al., 2010), it is possible that various peptides generated by food protein degradation exert bioactivity additively or synergistically and requires further investigation to improve the understanding.

1.4. Carbohydrates

Carbohydrates include a variety of molecules, which differ in constituent monosaccharides, degree of polymerization (DP), and types of glycosidic linkages in respective of their chemical structures. Carbohydrates in cow's milk are mostly low-molecular-weight carbohydrates and are dominated by lactose (4.4–5.8% in milk) (FoodData Central, 2022), with slight amounts of glucose (0.01%), galactose (0.01%), (Ohlsson et al., 2017) and oligosaccharides (<0.01%) (Fong, Ma, &

McJarrow, 2011; McJarrow & van Amelsfort-Schoonbeek, 2004). In comparison, plant materials contain carbohydrates encompassing a wide range of DP, from monosaccharides to polysaccharides with varied structures (Table 1.5). Nutritionally, food carbohydrates can be categorized as available and unavailable carbohydrates. Digestive enzymes can hydrolyze available carbohydrates to form monosaccharides, which the human intestine can absorb. The lack of specific glycolytic enzymes in the digestive tract or the inaccessibility of substrates makes certain carbohydrates unavailable to provide energy to the human body. Available carbohydrates include monosaccharides, most food-derived disaccharides, maltodextrin, and digestible starch. Accordingly, the rest of the carbohydrates in foods belong to unavailable carbohydrates, which include nondigestible oligosaccharides (NDO), non-starch polysaccharides (NSP), and resistant starch (RS). Although not directly providing energy to the human body, some of the unavailable carbohydrates display bioactivities that could be beneficial to human health.

The chemical composition of carbohydrates in plant-based beverages is not widely available in the literature. The carbohydrate composition data of plant materials may be used for roughly estimating their contents in the beverages. Nonetheless, the manufacturing process may chemically modify the molecular structure of some compounds in the plant materials as well as physically remove part of the components, rendering the compositions of beverage products considerably different from the raw plant materials and consequently increasing the uncertainty of the estimation.

The filtration step removes mainly insoluble materials with larger particle sizes in the slurry with ground plant materials. The raw material crops for making plant-based beverages are often rich in fibers, which may account for a significant portion of the insoluble materials being removed.

Table 1.5. Monosaccharides, disaccharides, oligosaccharides, and dietary fibers in plant materials for making beverages (g/ 100 g) and in cow's milk (g/ 100 g or g/ 100 mL).

Carbohydrates	Soybean	Almond	Rice		Oat[1]	Pea protein[2]	Cow's milk
			Brown	White			
Monosaccharides							
Glucose	0.030	0–0.87	0.023–0.212	0.072–0.315	0.06–0.1	0.00–0.31	0.01
Fructose	0.039	0–0.83	0.02–0.353	0.079–0.119	0–0.05		
Galactose		0–0.78					0.01
Disaccharides							
Sucrose	5.74–7.27	2.5–5.12	0.09–1.089	0.006–0.409	0.5–0.88	0.14–2.06	-
Maltose		0–0.14	0–0.89	0.010–0.044	0.01–0.03		-
Melibiose	0–0.09						-
Lactose	-	-	-	-	-	-	4.4–5.8
Oligosaccharides							
Raffinose	0.76–1.26	0.04–2.11	0–0.110	0.002–0.079	0.16–0.26	0.06–1.06	-
Stachyose	3.19–4.34				0.07–0.08	0.16–3.49	-
Verbascose	0.12–0.20					0.15–4.17	-
Manninotriose	0–0.13						-
Maltotriose			0–0.011	0–0.016			-
3'-Sialyllactose	-	-	-	-	-	-	0.0042–0.0055
6'-Sialyllactose	-	-	-	-	-	-	0.0004–0.0017
6'-Sialyllactosamine	-	-	-	-	-	-	0–0.0011
Dietary fiber							
Insoluble	9.22–32.6		1.38–4.10	0.8	7.67–9.24		-
Soluble	2.20–12.17		0.29–1.8	0.71	2.26–3.43		-
Total	13.6–35.9	7.9–19.3	2.5–7.1	0.2–1.51	8.8–18.6	0.1–2.5	-
β-glucan			0.1–0.39	0.11–0.71	2.7–6.3		-
Pectin	17						
Arabinoxylan			0.4–0.51		1.2–11.6		
Cellulose	8		0.3		0.6		
Starch	0.19–0.91	0.66–0.97	72–82	~90	34.8–61.0	0–8	-
Reference	(Bähr, Fechner,	(Barreira, Pereira,	(Bach Knudsen,	(Bao, 2019; FoodData	(Bach Knudsen et	(Bhatty & Christison,	(Fong et al., 2011; FoodData

	Hasenkopf, Mittermaier, & Jahreis, 2014; Choct, Dersjant-Li, McLeish, & Peisker, 2010; Fan, Zang, & Xing, 2015; Kawamura, Nagao, & Kasai, 1977; Kuo, VanMiddlesworth, & Wolf, 1988; Lee, Lee, Lee, Jang, & Kim, 2008; Písaříková & Zralý, 2010; Wilson, Birmingham, Moon, & Snyder, 1978)	Oliveira, & Ferreira, 2010; Cardozo & Li, 1994; FoodData Central, 2022; Gouta et al., 2020; Yada, Huang, & Lapsley, 2013)	Nørskov, Bolvig, Hedemann, & Laerke, 2017; Bao, 2019; Dodevska et al., 2013; FoodData Central, 2022; Lee et al., 2008; Mohan, Malleshi, & Koseki, 2010; Ramulu & Rao, 1997; Tran et al., 2005)	Central, 2022; Lee et al., 2008; Phuwadolpaisarn, 2021; Tran et al., 2005; Wang et al., 2020)	al., 2017; FoodData Central, 2022; Kuo et al., 1988; MacArthur & D'appolonia, 1979; Manthey, Hareland, & Huseby, 1999; Picolli da Silva & de Lourdes Santorio Ciocca, 2005; Tian, Gruppen, & Schols, 2015)	1984; Boukid, Rosell, & Castellari, 2021)	Central, 2022; McJarrow & van Amelsfort-Schoonbeek, 2004; Ohlsson et al., 2017)

[1] Data of whole, partially debranned, and debranned oat.

[2] Data of pea protein concentrate and isolate.

1.4.1. Mono- and disaccharides

Mono- and disaccharides, also known as simple sugars, present in foods are primarily available carbohydrates. Sucrose is the main simple sugar in most crops for making plant-based beverages (Table 1.5), including soybeans, almonds, oat, and pea (1.5–4.1 g/100 g in pea seeds (Bhatty & Christison, 1984; Fan et al., 2015; D. A. Jones, DuPont, Ambrose, Frias, & Hedley, 1999)). Monosaccharides, including glucose and fructose, are also present but in smaller amounts. These simple sugars are highly water-soluble and can be easily extracted and dissolved in water during the manufacturing process of plant-based beverages. Given the expected high recovery, the concentration of monosaccharides and disaccharides in beverage products is primarily determined by the concentration in plant materials and the dilution factor. Cultivars and growing and storage conditions could influence the contents of various carbohydrate molecules in plant materials. The isolation process of pea protein also plays an essential role in the carbohydrate composition (Bhatty & Christison, 1984; Boukid et al., 2021). Considering the concentration of simple sugars in the plant materials and the dilution factor, the amounts of simple sugars in unsweetened beverage products should be lower than in cow's milk. However, during the manufacturing process of western-style rice milk and oat milk, α-amylase and amyloglucosidase are often added for hydrolyzing starch to avoid the excessively high viscosity caused by starch gelatinization and increase the sweetness (Sethi, Tyagi, & Anurag, 2016; Silva, Silva, & Ribeiro, 2020). The enzymatic treatment can generate a significant amount of mono- and disaccharides (i.e., glucose and maltose) and maltodextrin. The glucose (<0.25–2.56 g/100 g) and maltose (<0.25–3.74 g/100 g) contents in commercial unsweetened oat milk products in the United States and the total sugar content (4.17–6.25 g/100 g) in unsweetened rice milk products (FoodData Central, 2022), which are generally much higher than the concentrations in oat and rice, reflecting the effect of starch

hydrolysis on the simple sugar contents. In the United States, mono- and disaccharides generated by such controlled enzymatic hydrolysis are labeled as added sugars per FDA guidance (U.S. Food and Drug Administration Center for Food Safety and Applied Nutrition, 2019). Aside from endogenous carbohydrates, exogenously added sugars might account for a majority of simple sugar in sweetened plant-based beverages. According to the nutrition label claims, the added sugar content is 2–7 g per 236 mL or 240 mL serving (0.8–3 g/100 mL) in various sweetened plant-based beverage products (almond, soy, and pea milk) in the United States.

1.4.2. Oligosaccharides

Digestible and nondigestible oligosaccharides may both be found in plant-based beverages. As mentioned above, western-style rice and oat milk may contain maltodextrin due to the hydrolysis of starch molecules with α-amylase. Maltodextrin is composed of glucose residues primarily linked by α-1,4 glycosidic bonds, with a smaller amount of α-1,6 glycosidic linkages. It can be rapidly digested after ingestion, resulting in an absorption rate close to the ingestion of pure glucose (Hofman, van Buul, & Brouns, 2016).

NDO in raw material crops are primarily raffinose family oligosaccharides (RFO) with DP of 3–5, namely raffinose, stachyose, and verbascose, respectively (Table 1.5). Their chemical structures include a sucrose core extended with one to three galactose residues from the glucose residue with α-1,6 glycosidic linkages. Legumes generally contain higher amounts of RFO than cereals and tree nuts. Pea seeds consist of 0.3–1.0 g/100 g raffinose, 0.7–3.8 g/100 g stachyose, and 0–3.1 g/100 g verbascose (Bhatty & Christison, 1984; Fan et al., 2015; D. A. Jones et al., 1999), which are comparable to the concentrations in soybeans. Similar to mono- and disaccharides, these oligosaccharides are also highly soluble in water and are expected to have a high recovery in beverage products. Nonetheless, pea protein isolation methods often differ among

manufacturers, leading to varied oligosaccharide contents in pea protein products (Bhatty & Christison, 1984; Boukid et al., 2021). The concentration of RFO was measured in commercial soy and almond milk products in the United States, with the summed concentration of RFO ranging from 0.430–0.640 g/100 g in soy milk and 0.0118–0.0194 g/100 g in almond milk (Huang, Paviani, Fukagawa, Phillips, & Barile, 2022). The commercial soy milk consisted of 0.047–0.071 g/100 g raffinose, 0.29–0.54 g/100 g stachyose, and 0.018–0.031 g/100 g verbascose (Huang et al., 2022); the relative abundance of the three oligosaccharides are in line with their concentrations in raw soybeans.

RFOs in legumes are often considered an undesirable factor because of their flatus-producing property; some processing methods, such as soaking and germination, have been studied for the purpose of RFO reduction (Jood, Mehta, Singh, & Bhat, 1985). However, RFO have also been reported displaying potential prebiotic activity. It was shown that raffinose could efficiently be utilized by *Bifidobacterium* species, such as *Bifidobacterium infantis*, *Bifidobacterium breve*, and *Bifidobacterium adolescentis*, and support their growth *in vitro* and *in vivo* (Amaretti et al., 2006). Dietary supplementation with soybean oligosaccharide extract was found to recover the abnormal lipid levels and oxidative stress caused in rats fed with a high-fat diet (Chen, Liu, Zhu, Xu, & Li, 2010) and increased the concentration of short-chain fatty acids (SCFAs) in the intestinal content of weaning mini-piglets (Zhou et al., 2014). However, a human study showed that 5 g/d of raffinose supplementation in healthy adults for three weeks did not significantly increase *Bifidobacterium* and *Lactobacillus* nor SCFAs in the feces (Fernando et al., 2010).

Oligosaccharides in cow's milk possess considerably different structures from plant-derived oligosaccharides. They have a lactose or lactosamine core, with additional monosaccharide residues extending the structures from the galactose residue (Robinson, 2019). 3'-Sialyllactose and

6’-sialyllactose are typically the most abundant oligosaccharides found in cow’s milk. Bovine milk oligosaccharides are present in liquid milk in a low abundance. Total oligosaccharide content in commercial soy milk (0.430–0.640 g/100 g) and almond milk (0.0118–0.0194 g/100 g) sold in the United States (Huang et al., 2022) is higher than that in cow’s milk (0.008–0.01 g/100 mL) (Durham, Cohen, Bunyatratchata, Fukagawa, & Barile, 2022). Some beneficial effects of bovine milk oligosaccharides were demonstrated by *in vivo* studies, such as modulating gut microbial composition, decreasing gut permeability, and reducing inflammation (Boudry et al., 2017; Hamilton et al., 2017; M. Wang et al., 2021); thus the oligosaccharides have the potential to be used as novel therapeutics. However, because the *in vivo* studies (Boudry et al., 2017; Hamilton et al., 2017; Wang et al., 2021) included high amounts of bovine milk oligosaccharides in the diet, oligosaccharides obtained from drinking cow’s milk not likely to provide similar functionalities.

1.4.3. Polysaccharides

Polysaccharides in plant materials for making plant-based beverages consist of various structures differing in monosaccharide composition, linkages, and DP. Due to the differences, they may play entirely different roles in human nutrition and health.

1.4.4. Starch

Starch consists of amylose and amylopectin and is present in high abundance in cereal endosperm. Amylose has linear structures composed of α-1,4-linked glucose residues; amylopectin is highly branched polysaccharides comprising glucose residues connected by α-1,4 linkages for the main and branch chains and α-1,6 linkages for the branch points. Starch accounts for 72–82% and ~90% of the total weight of brown and white rice, respectively. The high starch content in rice makes rice milk contain a high amount of available carbohydrates. The starch content in oats is lower (34.8–61.0%) than in rice and also varies with the degree of milling (Tian et al., 2015). The

amylose content varies greatly with rice varieties, ranging from 0% (waxy) to 31% (high-amylose) (Phuwadolpaisarn, 2021; Tao, Yu, Prakash, & Gilbert, 2019; Z. Zhou, Robards, Helliwell, & Blanchard, 2002), and is ~17–26% in oats (David M. Peterson, 1992). Pea seeds contain 39–44% starch (Bhatty & Christison, 1984; El-Adawy, Rahma, El-Bedawey, & El-Beltagy, 2003; Urbano et al., 2005); depending on the pea protein isolation process, starch could be completely or mostly removed. Soybeans and almonds usually only contain a meager amount of starch (<1%).

In the manufacturing process of plant-based beverages, the reduced material particle sizes and the addition of plenty of water aid in complete starch gelatinization during the heating steps. Hence, as mentioned earlier, starch molecules are usually hydrolyzed by α-amylase into dextrin, maltodextrin, and/or maltose during the production process of western-style rice milk and oat milk. The fully gelatinized starch and the hydrolysis products are both easily digestible and can be rapidly absorbed by the small intestine. For plant-based beverages made from starchy plant materials, this may, in turn, induce a high glycemic response, yet it was found that the digestion and absorption of starch are also affected by other food components, such as dietary fiber (e.g., arabinoxylans and β-glucans) and phenolics (Kim & White, 2013; Sajadimajd et al., 2019; Zhang, Dong, Hu, Ren, & Li, 2021).

1.4.5. Dietary fiber

Crops for making plant-based beverages contain varying amounts of dietary fiber (Table 1.5). In general, the total dietary fiber content in soybean (13.6–35.9 g/100 g), pea (9.1–22.0 g/100 g (Bähr et al., 2014; de Almeida Costa, da Silva Queiroz-Monici, Pissini Machado Reis, & de Oliveira, 2006; Martín-Cabrejas et al., 2003)), almond (7.9–19.3 g/100 g), and oat (8.8–18.6 g/100 g) is higher than in brown rice (3.29–7.1 g/100 g). White rice contains less dietary fiber (0.2–1.51 g/100 g) than brown rice due to the removal of the bran, which is the fiber-rich part of rice. Dietary

fiber in pea seeds can be removed during pea protein isolation; the content of dietary fiber in isolated pea protein usually decreases as the protein purity increases (Boukid et al., 2021). Based on water solubility, dietary fiber can be categorized as soluble and insoluble. Most plant materials, including soybean, rice, and oat, often contain more insoluble dietary fiber than soluble dietary fiber (Lee et al., 2008; Picolli da Silva & de Lourdes Santorio Ciocca, 2005; Písaříková & Zralý, 2010; Ramulu & Rao, 1997). The commonly applied filtration step in the manufacturing process of plant-based beverages might remove a significant portion of dietary fiber, especially the insoluble one.

Dietary fiber generally includes NDO, non-starch polysaccharides, resistant starch, and lignin (Jones, 2014). The plant materials for making plant-based beverages include a variety of non-starch polysaccharides, which will be discussed below.

1.4.5.1. β-Glucans

β-Glucans are polysaccharide molecules composed of β-linked glucose residues. Oat uniquely contains a substantial level of β-glucans (2.7–6.3 g/100 g) among the plant materials commonly used for making plant-based beverages. The β-glucan content in commercial oat milk products is still lacking; in an oat milk beverage produced in-house, with an oat-to-water ratio of 8:100 (w/v), the β-glucan concentration was determined to be 0.24 g/100 g (Bernat, Cháfer, González-Martínez, Rodríguez-García, & Chiralt, 2015). β-Glucans are also found in brown rice (0.1–0.39 g/100 g) and white rice (0.11–0.71 g/100 g) with lower abundances (Bach Knudsen et al., 2017; Dodevska et al., 2013; Phuwadolpaisarn, 2021) than in oats. In oat grains, β-glucans are more enriched in the bran fraction than the endosperm, likely due to the high abundance of starch in the endosperm (Westerlund, Andersson, & Åman, 1993). In β-glucans isolated from oat bran, the glucose residues are solely connected by β-1,3 and β-1,4 glycosidic linkages (Johansson et al.,

2000). In their structures, 1,4-linked cello-oligosaccharide units, typically comprising three to four glucose residues, were connected with β-1,3 glycosidic linkages (Johansson et al., 2000; Wang, Wood, Huang, & Cui, 2003).

The cholesterol- and postprandial blood glucose-lowering effects of β-glucans have been extensively studied. A meta-analysis of randomized controlled trials found that daily consumption of at least 3 g of oat β-glucans led to a significant reduction of low-density lipoprotein (LDL) and total cholesterol (Whitehead, Beck, Tosh, & Wolever, 2014). Another meta-analysis showed that oat β-glucans significantly reduced postprandial blood glucose and insulin levels (Zurbau, Noronha, Khan, Sievenpiper, & Wolever, 2021). It is noteworthy that several factors, including the conditions of oat growing, storage, and processing, may alter the physicochemical properties (e.g., molecular weight, solubility, and rheological properties) of oat β-glucans and could consequently affect the physiological activity (Grundy, Fardet, Tosh, Rich, & Wilde, 2018; Wang & Ellis, 2014; Wolever et al., 2010; Zurbau et al., 2021). It is believed that the increased viscosity of intestinal digesta caused by soluble polysaccharides, including β-glucans, and the consequently delayed food digestion and nutrient absorption are associated with the cholesterol- and blood glucose-lowering effects (Grundy et al., 2018; Qi Wang & Ellis, 2014). Other potential health benefits of oat β-glucans, such as modulating *in vitro* and *in vivo* immune functions (Estrada et al., 1997), decreasing *in vitro* human dermal cancer cell viability (Choromanska et al., 2015), and promoting propionate production during *in vitro* fecal fermentation (Carlson, Erickson, Hess, Gould, & Slavin, 2017), were also described.

1.4.5.2. Pectin

Pectin is a group of highly diverse and complex polysaccharides partially soluble in water. In general, pectin comprises a few different regions, including homogalacturonan (HG),

substituted homogalacturonan (e.g., rhamnogalacturonan II (RG II) and xylogalacturonan), and rhamnogalacturonan I (RG I) (Mohnen, 2008). In soybean and almond, pectin accounts for a great portion of non-starch polysaccharides, aside from cellulose (Choct et al., 2010; Dourado, Barros, Mota, Coimbra, & Gama, 2004). Pectin extracted from soybean is rich in galactose and arabinose, with a galacturonic acid to rhamnose ratio of around 8:1 (Brillouet & Carré, 1983). It was suggested that the isolated soybean pectin include both HG and RG I regions with neutral polysaccharides branches of DP = ~40 (e.g., arabinan) linked to the O-4 position of rhamnose residues (Brillouet & Carré, 1983). Soy pectin derived from okara (the insoluble residue resulting from defatted soybean through water extraction), namely soluble soybean polysaccharides (SSPS), which are also rich in galactose and arabinose, were revealed to exhibit potential prebiotic activities (Maeda & Nakamura, 2021; Min et al., 2015; Tian et al., 2016). The prebiotic property of SPSS was demonstrated by its ability to promote the production of short-chain fatty acids (SCFAs) *in vitro* and *in vivo* (Min et al., 2015; Tian et al., 2016). Similar to soybeans, pectin isolated from almond kernels also consists of neutral polysaccharide branches. The pectin structures were determined as highly-branched arabinan-rich pectin (Dourado, Madureira, et al., 2004). The structure of the arabinan branches was elucidated to be consisting of a backbone comprising α-1,5-linked arabinose residues with arabinose side chains connected with α-1,2 or α-1,3 glycosidic linkages (Dourado, Cardoso, Silva, Gama, & Coimbra, 2006; Dourado, Madureira, et al., 2004). Dourado, Madureira, et al. (2004) found that pectin derived from almond kernels exhibited *in vitro* immunomodulatory activity in murine cells.

1.4.5.3. Arabinoxylans

Arabinoxylans are commonly present in cereals, including rice and oat. Their structures consist of a β-1,4 xylopyranosyl backbone substituted by arabinose residues on O-2 and/or O-3

positions; the substitution also could be other monosaccharides (e.g., glucuronic acid) and disaccharides (e.g., arabinobiose and xylosyl-arabinose) (Pastell, Virkki, Harju, Tuomainen, & Tenkanen, 2009; Shibuya, 1989; Westerlund et al., 1993). The degree of substitution, the substituting position, and the side chain carbohydrate composition vary with the location in cereal grains and the type of cereal. For example, for both oat and rice, arabinoxylan in the bran layer contains a higher ratio of disubstituted (i.e., O-2 and O-3) to monosubstituted (i.e., O-2 or O-3) xylose residues on the backbone than in the endosperm (Shibuya, 1989; Westerlund et al., 1993). Also, arabinose substitution in oat arabinoxylans are highly clustered, as opposed to the more randomly distributed substitution found in arabinoxylans of other cereal sources (Tian et al., 2015). In addition to carbohydrate moieties, phenolic acids, such as ferulic acid, *p*-coumaric acid, and diferulic acid, are often esterified to terminal arabinose residues of arabinoxylan (Bunzel, Ralph, Marita, Hatfield, & Steinhart, 2001; Shibuya, 1984; J. Wang et al., 2020). The phenolic acid esters can cross-link between arabinoxylans or associate arabinoxylans with other molecules in cereals, e.g., proteins and lignin (Bunzel et al., 2001; Wang et al., 2020).

The bioactivities of arabinoxylans from various cereal sources have been studied, including antioxidant activity (due to the presence of phenolic acids), prebiotic activity, and reduction of postprandial blood glucose (possibly due to the high viscosity of arabinoxylans with specific structural properties, similar to β-glucans) (He et al., 2021).

The prebiotic effects are primarily conducted with arabinoxylan and arabinoxylan-oligosaccharides derived from other cereal sources, such as wheat bran and corn cobs (Broekaert et al., 2011). Some *in vitro* studies revealed that the fermentability of arabinoxylan is associated with its water solubility as well as the degree and types of substitution on the xylan backbone (Broekaert et al., 2011; Kabel, Kortenoeven, Schols, & Voragen, 2002; Karppinen, Kiiliäinen,

Liukkonen, Forssell, & Poutanen, 2001). Karppinen et al. (2001) showed that water-exactable arabinoxylans of rye bran were fermented faster than the water-unextractable fraction in a human fecal inoculum. In the same study, the authors also found that arabinoxylans with higher degrees of arabinose substitution (Ara/Xyl ratio > 1) were more difficult to utilize by fecal bacteria. The bifidogenic activity of arabionoxylo-oligosaccharides and arabinoxylans was demonstrated in several *in vitro* and *in vivo* studies (Broekaert et al., 2011; Neyrinck et al., 2011). For example, in an *in vitro* study using human fecal microbiota, three wheat bran-derived arabinoxylan fractions with average molecular weights of 66, 278, and 354 kDa resulted in selective growth of the *Bifidobacterium* and a significant increase of SCFAs (Hughes et al., 2007). Similar effects were also observed in *in vivo* studies using rodents and humans with arabinoxylan or arabinoxylan-oligosaccharides, with a wide range of average DP (average DP 3–284) (Cloetens et al., 2010; Damen et al., 2011; Neyrinck et al., 2011; Van Craeyveld et al., 2008; Van den Abbeele et al., 2011).

Unlike β-glucans, which are relatively water-soluble, arabinoxylans in cereal grains is primarily insoluble in water (He et al., 2021; Tian et al., 2015; Westerlund et al., 1993). For example, Tian et al. (2015) showed that whole grain oats contained 11.6 g/100g of arabinoxylans, but only 0.9 g/100 g was water-soluble. The low water solubility may greatly limit the proportion of arabinoxylans remaining in the plant-based beverage products when the manufacturing process includes a filtration step. The water solubility of arabinoxylans may also vary with the processing techniques being applied. Some process methods were demonstrated to increase the water-extractable arabinoxylans or produce arabinoxylan-oligosaccharides, such as enzymatic hydrolysis, microwave-assisted autohydrolysis, high-pressure treatment, and ball-milling (Kim et al., 2015; Lai & Huang, 2014; Rose & Inglett, 2010; Truong & Rumpagaporn, 2019). Given the

potential health benefits of arabinoxylans, more studies are needed to explore the strategies to aid the solubilization of arabinoxylans and the generation of arabinoxylan-oligosaccharides in the manufacturing process of plant-based beverages.

1.5. Summary and future perspectives

Plant-based beverages have become a fast-growing segment in new food product development. Consumers are also demanding more product diversification and showing a greater interest in other plant-based dairy alternatives such as yogurts, ice creams, and other plant-based desserts, although this trend is still modest in comparison with the market share of plant-based beverages.

Despite the similar appearance and texture, the composition and nutritional value of plant-based beverages and cow's milk are fundamentally different. Most commercial plant-based beverages have much lower protein quantity and quality than cow's milk, except for soy milk and some, but not all, pea milk products. Unlike cow's milk, which can provide sufficient essential amino acids for all consumers except for infants, most plant-based beverages lack specific essential amino acids, except for soy milk, which is a source of complete protein for consumers above three years of age. Some plant-based beverages contain proteinaceous bioactive compounds, such as lectins, BBIs, and lunasin. The bioactivities of these compounds are multifaceted and sometimes display both beneficial and detrimental properties. More studies are needed to elucidate their influence on human health.

Unsweetened plant-based beverages made from soybeans, almonds, and peas contain lower levels of available carbohydrates than cow's milk, whereas cereal-based beverages usually contain higher levels of available carbohydrates due to the high starch content in cereal grains. Varying

levels of dietary fiber with diverse structures are present in plant-based beverages. Some of them possess potential bioactivities, such as the prebiotic activities of RFO and arabinoxylan-oligosaccharides and the cholesterol-lowering property of β-glucans.

Different approaches that have the potential to improve plant-based beverages' nutritional values and bioactive properties demand more investigation. For example, blending various crops, including conventional and new plant materials, that complement the limiting amino acid in the other crops could improve overall protein quality in beverage products. Unconventional processing techniques, such as high-pressure processing, media milling, and enzymatic treatments using proteolytic and glycolytic enzymes, could aid in increasing shelf-life and improving protein and carbohydrates solubility, and potentially generate bioactive components such as low-molecular-weight peptides and oligosaccharides derived from polysaccharides. The allergenicity of certain plant-based beverages may also be reduced through specific processing.

In vivo bioactivities are highly related to the dose and bioavailability. However, the information about the levels of many bioactive compounds in plant-based beverages is still scarcely available. The composition and concentration of bioactive components in plant-based beverages profoundly depend on the properties of raw materials and the processing methods. Characterizing various bioactive molecules in plant-based beverages already available on the market and those still undergoing development, with the use of new materials or innovative processing methods, will help understand their potential bioactivities. Studying the bioavailability of the bioactive components is also crucial to evaluating their potential influence on human health. For conducting these researches, robust and validated analytical methods are necessary. Analytical techniques that allow identifying and quantifying specific compounds, such as chromatography,

mass spectrometry, and nuclear magnetic resonance, can be applied in these studies for a better understanding of plant-based beverages' bioactive properties.

Lastly, plant-based beverages' chemical composition, nutritional values, and potential health benefits differ considerably among different products. Although Nutrition Facts on commercial beverages' packages include the general composition, they are often derived from food composition databases and do not inform about the specific protein quality and detailed carbohydrate composition. More comprehensive information about the composition and the ratio of plant materials added to the ingredients should be made transparently available to the consumers. Considering that the current regulatory structure is not ready to deal with novel protein sources, there is an opportunity to start from ground zero and identify new communication channels that will be effective for the industry and ensure consumer safety and awareness. These changes will require government, scientists, and industry regulators to select quality measurements to safeguard transparency, maintain a high standard for science, and yet deliver accurate descriptors that consumers understand.

References

Abelilla, J. J., Liu, Y., & Stein, H. H. (2018). Digestible indispensable amino acid score (DIAAS) and protein digestibility corrected amino acid score (PDCAAS) in oat protein concentrate measured in 20- to 30-kilogram pigs. Journal of the science of food and agriculture, 98(1), 410–414. https://doi.org/10.1002/jsfa.8457

Alaswad, A. A., & Krishnan, H. B. (2016). Immunological investigation for the presence of lunasin, a chemopreventive soybean peptide, in the seeds of diverse plants. Journal of Agricultural and Food Chemistry, 64(14), 2901–2909. https://doi.org/10.1021/acs.jafc.6b00445

Albillos, S. M., Jin, T., Howard, A., Zhang, Y., Kothary, M. H., & Fu, T.-J. (2008). Purification, crystallization and preliminary X-ray characterization of prunin-1, a major component of the almond (Prunus dulcis) allergen amandin. Journal of Agricultural and Food Chemistry, 56(13), 5352–5358. https://doi.org/10.1021/jf800529k

Almond Breeze. (2022). Unsweetened - Almond Breeze. Retrieved May 9, 2022, from https://www.almondbreeze.co.uk/unsweetened/

Alpro. (2022). Alpro. Retrieved May 8, 2022, from https://www.alpro.com/uk/

Amaretti, A., Tamburini, E., Bernardi, T., Pompei, A., Zanoni, S., Vaccari, G., … Rossi, M. (2006). Substrate preference of Bifidobacterium adolescentis MB 239: compared growth on single and mixed carbohydrates. Applied Microbiology and Biotechnology, 73(3), 654–662. https://doi.org/10.1007/s00253-006-0500-9

Amino, Y., Nakazawa, M., Kaneko, M., Miyaki, T., Miyamura, N., Maruyama, Y., & Eto, Y. (2016). Structure-CaSR-Activity Relation of Kokumi γ-Glutamyl Peptides. Chemical & Pharmaceutical Bulletin, 64(8), 1181–1189. https://doi.org/10.1248/cpb.c16-00293

Armour, J. C., Perera, R. L. C., Buchan, W. C., & Grant, G. (1998). Protease inhibitors and lectins in soya beans and effects of aqueous heat-treatment. Journal of the Science of Food and Agriculture.

Arques, M. C., Marín-Manzano, M. C., da Rocha, L. C. B., Hernandez-Ledesma, B., Recio, I., & Clemente, A. (2014). Quantitative determination of active Bowman-Birk isoinhibitors, IBB1 and IBBD2, in commercial soymilks. Food chemistry, 155, 24–30. https://doi.org/10.1016/j.foodchem.2014.01.024

Arques, M. C., Pastoriza, S., Delgado-Andrade, C., Clemente, A., & Rufián-Henares, J. A. (2016). Relationship between Glycation and Polyphenol Content and the Bioactivity of Selected Commercial Soy Milks. Journal of Agricultural and Food Chemistry, 64(8), 1823–1830. https://doi.org/10.1021/acs.jafc.6b00181

Aydar, E. F., Tutuncu, S., & Ozcelik, B. (2020). Plant-based milk substitutes: Bioactive compounds, conventional and novel processes, bioavailability studies, and health effects. Journal of functional foods, 70, 103975. https://doi.org/10.1016/j.jff.2020.103975

Bach Knudsen, K. E., Nørskov, N. P., Bolvig, A. K., Hedemann, M. S., & Laerke, H. N. (2017). Dietary fibers and associated phytochemicals in cereals. Molecular Nutrition & Food Research, 61(7). https://doi.org/10.1002/mnfr.201600518

Baek, J. M., & Kim, S. I. (1993). Nucleotide sequence of a cDNA encoding soybean Bowman-Birk proteinase inhibitor. Plant Physiology, 102(2), 687. https://doi.org/10.1104/pp.102.2.687

Baek, J. M., Song, J. C., Choi, Y. D., & Kim, S. I. (1994). Nucleotide sequence homology of cDNAs encoding soybean Bowman-Birk type proteinase inhibitor and its isoinhibitors. Bioscience, Biotechnology, and Biochemistry, 58(5), 843–846. https://doi.org/10.1271/bbb.58.843

Bahna, S. L. (2008). Hypoallergenic formulas: optimal choices for treatment versus prevention. Annals of Allergy, Asthma & Immunology, 101(5), 453–9; quiz 459. https://doi.org/10.1016/S1081-1206(10)60281-5

Bähr, M., Fechner, A., Hasenkopf, K., Mittermaier, S., & Jahreis, G. (2014). Chemical composition of dehulled seeds of selected lupin cultivars in comparison to pea and soya bean. LWT - Food Science and Technology, 59(1), 587–590. https://doi.org/10.1016/j.lwt.2014.05.026

Bao, J. (2019). Rice starch. In Rice (pp. 55–108). Elsevier. https://doi.org/10.1016/B978-0-12-811508-4.00003-4

Barac, M., Cabrilo, S., Pesic, M., Stanojevic, S., Zilic, S., Macej, O., & Ristic, N. (2010). Profile and functional properties of seed proteins from six pea (Pisum sativum) genotypes. International Journal of Molecular Sciences, 11(12), 4973–4990. https://doi.org/10.3390/ijms11124973

Barreira, J. C. M., Pereira, J. A., Oliveira, M. B. P. P., & Ferreira, I. C. F. R. (2010). Sugars profiles of different chestnut (Castanea sativa Mill.) and almond (Prunus dulcis) cultivars by HPLC-RI. Plant foods for human nutrition (Dordrecht, Netherlands), 65(1), 38–43. https://doi.org/10.1007/s11130-009-0147-7

Benedict, M. (2006). Food allergies: new labeling guidelines issued by the FDA. Health Care Food & Nutrition Focus, 23(8), 7.

Bernat, N., Cháfer, M., González-Martínez, C., Rodríguez-García, J., & Chiralt, A. (2015). Optimisation of oat milk formulation to obtain fermented derivatives by using probiotic Lactobacillus reuteri microorganisms. Food science and technology international = Ciencia y tecnologia de los alimentos internacional, 21(2), 145–157. https://doi.org/10.1177/1082013213518936

Bhatty, R. S., & Christison, G. I. (1984). Composition and nutritional quality of pea (Pisum sativum L.), faba bean (Vicia faba L. spp. minor) and lentil (Lens culinaris Medik.) meals, protein concentrates and isolates. Plant Foods for Human Nutrition, 34(1), 41–51. https://doi.org/10.1007/BF01095071

Blow, D. M., Janin, J., & Sweet, R. M. (1974). Mode of action of soybean trypsin inhibitor (Kunitz) as a model for specific protein-protein interactions. Nature, 249(452), 54–57. https://doi.org/10.1038/249054a0

Bösterling, B., & Quast, U. (1981). Soybean trypsin inhibitor (Kunitz) is doubleheaded. Biochimica et Biophysica Acta (BBA) - Enzymology, 657(1), 58–72. https://doi.org/10.1016/0005-2744(81)90130-3

Boudry, G., Hamilton, M. K., Chichlowski, M., Wickramasinghe, S., Barile, D., Kalanetra, K. M., … Raybould, H. E. (2017). Bovine milk oligosaccharides decrease gut permeability and improve inflammation and microbial dysbiosis in diet-induced obese mice. Journal of Dairy Science, 100(4), 2471–2481. https://doi.org/10.3168/jds.2016-11890

Boukid, F., Rosell, C. M., & Castellari, M. (2021). Pea protein ingredients: A mainstream ingredient to (re)formulate innovative foods and beverages. Trends in food science & technology, 110, 729–742. https://doi.org/10.1016/j.tifs.2021.02.040

Bradbury, J. H., Hammer, B. C., & Sugani, I. (1992). Heat stability of trypsin inhibitors in tropical root crops and rice and its significance for nutrition. Journal of the science of food and agriculture, 58(1), 95–100. https://doi.org/10.1002/jsfa.2740580116

Brillouet, J.-M., & Carré, B. (1983). Composition of cell walls from cotyledons of Pisum sativum, Vicia faba and Glycine max. Phytochemistry, 22(4), 841–847. https://doi.org/10.1016/0031-9422(83)85009-2

Broekaert, W. F., Courtin, C. M., Verbeke, K., Van deaaa 1Wiele, T., Verstraete, W., & Delcour, J. A. (2011). Prebiotic and other health-related effects of cereal-derived arabinoxylans, arabinoxylan-oligosaccharides, and xylooligosaccharides. Critical reviews in food science and nutrition, 51(2), 178–194. https://doi.org/10.1080/10408390903044768

Brosnan, J. T., & Brosnan, M. E. (2006). The sulfur-containing amino acids: an overview. The Journal of Nutrition, 136(6 Suppl), 1636S–1640S. https://doi.org/10.1093/jn/136.6.1636S

Bunzel, M., Ralph, J., Marita, J. M., Hatfield, R. D., & Steinhart, H. (2001). Diferulates as structural components in soluble and insoluble cereal dietary fibre. Journal of the science of food and agriculture, 81(7), 653–660. https://doi.org/10.1002/jsfa.861

Capriotti, A. L., Caruso, G., Cavaliere, C., Samperi, R., Ventura, S., Zenezini Chiozzi, R., & Laganà, A. (2015). Identification of potential bioactive peptides generated by simulated gastrointestinal digestion of soybean seeds and soy milk proteins. Journal of Food Composition and Analysis, 44, 205–213. https://doi.org/10.1016/j.jfca.2015.08.007

Cardozo, M. S., & Li, B. W. (1994). Total dietary fiber content of selected nuts by two enzymatic–gravimetric methods. Journal of Food Composition and Analysis, 7(1-2), 37–43. https://doi.org/10.1006/jfca.1994.1004

Carlson, J. L., Erickson, J. M., Hess, J. M., Gould, T. J., & Slavin, J. L. (2017). Prebiotic Dietary Fiber and Gut Health: Comparing the in Vitro Fermentations of Beta-Glucan, Inulin and Xylooligosaccharide. Nutrients, 9(12). https://doi.org/10.3390/nu9121361

Cavazos, A., Morales, E., Dia, V. P., & De Mejia, E. G. (2012). Analysis of lunasin in commercial and pilot plant produced soybean products and an improved method of lunasin purification. Journal of Food Science, 77(5), C539–45. https://doi.org/10.1111/j.1750-3841.2012.02676.x

Chanson-Rolle, A., Aubin, F., Braesco, V., Hamasaki, T., & Kitakaze, M. (2015). Influence of the Lactotripeptides Isoleucine-Proline-Proline and Valine-Proline-Proline on Systolic Blood Pressure in Japanese Subjects: A Systematic Review and Meta-Analysis of Randomized Controlled Trials. Plos One, 10(11), e0142235. https://doi.org/10.1371/journal.pone.0142235

Chatterjee, C., Gleddie, S., & Xiao, C.-W. (2018). Soybean bioactive peptides and their functional properties. Nutrients, 10(9). https://doi.org/10.3390/nu10091211

Chen, H., Liu, L., Zhu, J., Xu, B., & Li, R. (2010). Effect of soybean oligosaccharides on blood lipid, glucose levels and antioxidant enzymes activity in high fat rats. Food chemistry, 119(4), 1633–1636. https://doi.org/10.1016/j.foodchem.2009.09.056

Chen, J., Mao, S., Xie, Y., Cao, Z., Zhang, Y., Liu, J., … Gu, H. (2006). Expression and inhibitory activity analysis of a 25-kD Bowman-Birk protease inhibitor in rice. Chinese Science Bulletin, 51(1), 54–62. https://doi.org/10.1007/s11434-005-0937-8

Choct, M., Dersjant-Li, Y., McLeish, J., & Peisker, M. (2010). Soy Oligosaccharides and Soluble Non-starch Polysaccharides: A Review of Digestion, Nutritive and Anti-nutritive Effects in Pigs and Poultry. Asian-Australasian journal of animal sciences, 23(10), 1386–1398. https://doi.org/10.5713/ajas.2010.90222

Choromanska, A., Kulbacka, J., Rembialkowska, N., Pilat, J., Oledzki, R., Harasym, J., & Saczko, J. (2015). Anticancer properties of low molecular weight oat beta-glucan – An in vitro study. International journal of biological macromolecules, 80, 23–28. https://doi.org/10.1016/j.ijbiomac.2015.05.035

Clemente, A., Carmen Marín-Manzano, M., Jiménez, E., Carmen Arques, M., & Domoney, C. (2012). The anti-proliferative effect of TI1B, a major Bowman-Birk isoinhibitor from pea (Pisum sativum L.), on HT29 colon cancer cells is mediated through protease inhibition. The British Journal of Nutrition, 108 Suppl 1, S135–44. https://doi.org/10.1017/S000711451200075X

Clemente, A., Gee, J. M., Johnson, I. T., Mackenzie, D. A., & Domoney, C. (2005). Pea (Pisum sativum L.) protease inhibitors from the Bowman-Birk class influence the growth of human colorectal adenocarcinoma HT29 cells in vitro. Journal of Agricultural and Food Chemistry, 53(23), 8979–8986. https://doi.org/10.1021/jf051528w

Clemente, A., Moreno, F. J., Marín-Manzano, M. del C., Jiménez, E., & Domoney, C. (2010). The cytotoxic effect of Bowman-Birk isoinhibitors, IBB1 and IBBD2, from soybean (Glycine max) on HT29 human colorectal cancer cells is related to their intrinsic ability to inhibit serine proteases. Molecular Nutrition & Food Research, 54(3), 396–405. https://doi.org/10.1002/mnfr.200900122

Cloetens, L., Broekaert, W. F., Delaedt, Y., Ollevier, F., Courtin, C. M., Delcour, J. A., … Verbeke, K. (2010). Tolerance of arabinoxylan-oligosaccharides and their prebiotic activity in healthy subjects: a randomised, placebo-controlled cross-over study. The British Journal of Nutrition, 103(5), 703–713. https://doi.org/10.1017/S0007114509992248

Craddock, J. C., Genoni, A., Strutt, E. F., & Goldman, D. M. (2021). Limitations with the Digestible Indispensable Amino Acid Score (DIAAS) with Special Attention to Plant-Based Diets: a Review. Current nutrition reports, 10(1), 93–98. https://doi.org/10.1007/s13668-020-00348-8

Cruz-Huerta, E., Fernández-Tomé, S., Arques, M. C., Amigo, L., Recio, I., Clemente, A., & Hernández-Ledesma, B. (2015). The protective role of the Bowman-Birk protease inhibitor in soybean lunasin digestion: the effect of released peptides on colon cancer growth. Food & function, 6(8), 2626–2635. https://doi.org/10.1039/c5fo00454c

Cui, L., Bandillo, N., Wang, Y., Ohm, J.-B., Chen, B., & Rao, J. (2020). Functionality and structure of yellow pea protein isolate as affected by cultivars and extraction pH. Food hydrocolloids, 108, 106008. https://doi.org/10.1016/j.foodhyd.2020.106008

Daliri, E. B.-M., Ofosu, F. K., Chelliah, R., Park, M. H., Kim, J.-H., & Oh, D.-H. (2019). Development of a Soy Protein Hydrolysate with an Antihypertensive Effect. International Journal of Molecular Sciences, 20(6). https://doi.org/10.3390/ijms20061496

Daliri, E. B.-M., Oh, D. H., & Lee, B. H. (2017). Bioactive Peptides. Foods, 6(5). https://doi.org/10.3390/foods6050032

Damen, B., Verspreet, J., Pollet, A., Broekaert, W. F., Delcour, J. A., & Courtin, C. M. (2011). Prebiotic effects and intestinal fermentation of cereal arabinoxylans and arabinoxylan oligosaccharides in rats depend strongly on their structural properties and joint presence. Molecular Nutrition & Food Research, 55(12), 1862–1874. https://doi.org/10.1002/mnfr.201100377

de Almeida Costa, G. E., da Silva Queiroz-Monici, K., Pissini Machado Reis, S. M., & 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 la Barca, A. M. C., Vázquez-Moreno, L., & Robles-Burgueño, M. R. (1991). Active soybean lectin in foods: Isolation and quantitation. Food chemistry, 39(3), 321–327. https://doi.org/10.1016/0308-8146(91)90149-I

de Mejia, E. G., Bradford, T., & Hasler, C. (2003). The anticarcinogenic potential of soybean lectin and lunasin. Nutrition Reviews, 61(7), 239–246. https://doi.org/10.1301/nr.2003.jul.239-246

Dhakal, S., Liu, C., Zhang, Y., Roux, K. H., Sathe, S. K., & Balasubramaniam, V. M. (2014). Effect of high pressure processing on the immunoreactivity of almond milk. Food Research International, 62, 215–222. https://doi.org/10.1016/j.foodres.2014.02.021

Dia, V. P., Torres, S., De Lumen, B. O., Erdman, J. W., & De Mejia, E. G. (2009). Presence of lunasin in plasma of men after soy protein consumption. Journal of Agricultural and Food Chemistry, 57(4), 1260–1266. https://doi.org/10.1021/jf803303k

Dinelli, G., Bregola, V., Bosi, S., Fiori, J., Gotti, R., Simonetti, E., … Hrelia, S. (2014). Lunasin in wheat: a chemical and molecular study on its presence or absence. Food chemistry, 151, 520–525. https://doi.org/10.1016/j.foodchem.2013.11.119

Dodevska, M. S., Djordjevic, B. I., Sobajic, S. S., Miletic, I. D., Djordjevic, P. B., & Dimitrijevic-Sreckovic, V. S. (2013). Characterisation of dietary fibre components in cereals and legumes used in Serbian diet. Food chemistry, 141(3), 1624–1629. https://doi.org/10.1016/j.foodchem.2013.05.078

Domoney, C., Welham, T., Sidebottom, C., & Firmin, J. L. (1995). Multiple isoforms of Pisum trypsin inhibitors result from modification of two primary gene products. FEBS Letters, 360(1), 15–20. https://doi.org/10.1016/0014-5793(95)00070-p

Dourado, F., Barros, A., Mota, M., Coimbra, M. A., & Gama, F. M. (2004). Anatomy and cell wall polysaccharides of almond (Prunus dulcis D. A. Webb) seeds. Journal of Agricultural and Food Chemistry, 52(5), 1364–1370. https://doi.org/10.1021/jf030061r

Dourado, F., Cardoso, S. M., Silva, A. M. S., Gama, F. M., & Coimbra, M. A. (2006). NMR structural elucidation of the arabinan from Prunus dulcis immunobiological active pectic polysaccharides. Carbohydrate Polymers, 66, 27–33.

Dourado, F., Madureira, P., Carvalho, V., Coelho, R., Coimbra, M. A., Vilanova, M., … Gama, F. M. (2004). Purification, structure and immunobiological activity of an arabinan-rich pectic polysaccharide from the cell walls of Prunus dulcis seeds. Carbohydrate Research, 339, 2555–2566.

Durham, S. D., Cohen, J. L., Bunyatratchata, A., Fukagawa, N. K., & Barile, D. (2022). Oligosaccharides. In Encyclopedia of dairy sciences (pp. 141–153). Elsevier. https://doi.org/10.1016/B978-0-12-818766-1.00182-3

El-Adawy, T. A., Rahma, E. H., El-Bedawey, A. A., & El-Beltagy, A. E. (2003). Nutritional potential and functional properties of germinated mung bean, pea and lentil seeds. Plant foods for human nutrition (Dordrecht, Netherlands), 58(3), 1–13. https://doi.org/10.1023/B:QUAL.0000040339.48521.75

Eriksen, S. (1983). Application of enzymes in soy milk production to improve yield. Journal of Food Science, 48(2), 445–447. https://doi.org/10.1111/j.1365-2621.1983.tb10762.x

Estrada, A., Yun, C. H., Van Kessel, A., Li, B., Hauta, S., & Laarveld, B. (1997). Immunomodulatory activities of oat beta-glucan in vitro and in vivo. Microbiology and Immunology, 41(12), 991–998. https://doi.org/10.1111/j.1348-0421.1997.tb01959.x

Eyarkai Nambi, V., Manickavasagan, A., & Shahir, S. (2017). Rice milling technology to produce brown rice. In A. Manickavasagan, C. Santhakumar, & N. Venkatachalapathy (eds.), Brown Rice (pp. 3–21). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-59011-0_1

Fan, P.-H., Zang, M.-T., & Xing, J. (2015). Oligosaccharides composition in eight food legumes species as detected by high-resolution mass spectrometry. Journal of the science of food and agriculture, 95(11), 2228–2236. https://doi.org/10.1002/jsfa.6940

Fekete, Á. A., Givens, D. I., & Lovegrove, J. A. (2015). Casein-derived lactotripeptides reduce systolic and diastolic blood pressure in a meta-analysis of randomised clinical trials. Nutrients, 7(1), 659–681. https://doi.org/10.3390/nu7010659

Fernández-Tomé, S., Sanchón, J., Recio, I., & Hernández-Ledesma, B. (2018). Transepithelial transport of lunasin and derived peptides: Inhibitory effects on the gastrointestinal cancer cells viability. Journal of Food Composition and Analysis, 68, 101–110. https://doi.org/10.1016/j.jfca.2017.01.011

Fernando, W. M. U., Hill, J. E., Zello, G. A., Tyler, R. T., Dahl, W. J., & Van Kessel, A. G. (2010). Diets supplemented with chickpea or its main oligosaccharide component raffinose modify faecal microbial composition in healthy adults. Beneficial microbes, 1(2), 197–207. https://doi.org/10.3920/BM2009.0027

Ferrasson, E., Quillien, L., & Gueguen, J. (1995). Amino acid sequence of a Bowman-Birk proteinase inhibitor from pea seeds. Journal of protein chemistry, 14(6), 467–475. https://doi.org/10.1007/BF01888141

Foltz, M., van der Pijl, P. C., & Duchateau, G. S. M. J. E. (2010). Current in vitro testing of bioactive peptides is not valuable. The Journal of Nutrition, 140(1), 117–118. https://doi.org/10.3945/jn.109.116228

Fong, B., Ma, K., & McJarrow, P. (2011). Quantification of bovine milk oligosaccharides using liquid chromatography-selected reaction monitoring-mass spectrometry. Journal of Agricultural and Food Chemistry, 59(18), 9788–9795. https://doi.org/10.1021/jf202035m

Food and Agriculture Organization of the United Nations/World Health Organization. (2013). Dietary protein quality evaluation in human nutrition - Report of an' ' FAO Expert Consultation (No. FAO FOOD AND NUTRITION PAPER 92). Rome, Italy: Food and Agriculture Organization of the United Nations.

FoodData Central. (2022). FoodData Central. Retrieved July 7, 2022, from https://fdc.nal.usda.gov/index.html

Fujiwara, T., Hirai, M. Y., Chino, M., Komeda, Y., & Naito, S. (1992). Effects of sulfur nutrition on expression of the soybean seed storage protein genes in transgenic petunia. Plant Physiology, 99(1), 263–268. https://doi.org/10.1104/pp.99.1.263

Galvez, A. F., & de Lumen, B. O. (1999). A soybean cDNA encoding a chromatin-binding peptide inhibits mitosis of mammalian cells. Nature Biotechnology, 17(5), 495–500. https://doi.org/10.1038/8676

Gatehouse, J. A., Croy, R. R., Morton, H., Tyler, M., & Boulter, D. (1981). Characterisation and subunit structures of the vicilin storage proteins of pea (Pisum sativum L.). European Journal of Biochemistry / FEBS, 118(3), 627–633. https://doi.org/10.1111/j.1432-1033.1981.tb05565.x

Gitlin-Domagalska, A., Maciejewska, A., & Dębowski, D. (2020). Bowman-Birk Inhibitors: Insights into Family of Multifunctional Proteins and Peptides with Potential Therapeutical Applications. Pharmaceuticals (Basel, Switzerland), 13(12). https://doi.org/10.3390/ph13120421

Gouta, H., Ksia, E., Laaribi, I., Molino Gahete, F., Estopañan, G., Juan, T., … Martínez-García, P. J. (2020). Evaluation of the chemical and nutritional properties of Tunisian almond cultivars. Italian Journal of Food Science.

Grand View Research. (2020). Plant-based Beverages Market Size Report, 2021-2028. Retrieved June 10, 2022, from https://www.grandviewresearch.com/industry-analysis/plant-based-beverages-market

Grant, G., Alonso, R., Edwards, J. E., & Murray, S. (2000). Dietary soya beans and kidney beans stimulate secretion of cholecystokinin and pancreatic digestive enzymes in 400-day-old Hooded-Lister rats but only soya beans induce growth of the pancreas. Pancreas, 20(3), 305–312. https://doi.org/10.1097/00006676-200004000-00013

Grant, G., Dorward, P. M., & Pusztai, A. (1993). Pancreatic enlargement is evident in rats fed diets containing raw soybeans (Glycine max) or cowpeas (Vigna unguiculata) for 800 days but not in those fed diets based on kidney beans (Phaseolus vulgaris) or lupinseed (Lupinus angustifolius). The Journal of Nutrition, 123(12), 2207–2215. https://doi.org/10.1093/jn/123.12.2207

Grundy, M. M.-L., Fardet, A., Tosh, S. M., Rich, G. T., & Wilde, P. J. (2018). Processing of oat: the impact on oat's cholesterol lowering effect. Food & function, 9(3), 1328–1343. https://doi.org/10.1039/c7fo02006f

Gueguen, J., & Barbot, J. (1988). Quantitative and qualitative variability of pea (Pisum sativum L.) protein composition. Journal of the science of food and agriculture, 42(3), 209–224. https://doi.org/10.1002/jsfa.2740420304

Guha, S., & Majumder, K. (2022). Comprehensive Review of γ-Glutamyl Peptides (γ-GPs) and Their Effect on Inflammation Concerning Cardiovascular Health. Journal of Agricultural and Food Chemistry. https://doi.org/10.1021/acs.jafc.2c01712

Guo, H. Y., Pang, K., Zhang, X. Y., Zhao, L., Chen, S. W., Dong, M. L., & Ren, F. Z. (2007). Composition, physiochemical properties, nitrogen fraction distribution, and amino acid profile of donkey milk. Journal of Dairy Science, 90(4), 1635–1643. https://doi.org/10.3168/jds.2006-600

Habiba, R. A. (2002). Changes in anti-nutrients, protein solubility, digestibility, and HCl-extractability of ash and phosphorus in vegetable peas as affected by cooking methods. Food chemistry, 77(2), 187–192. https://doi.org/10.1016/S0308-8146(01)00335-1

Hamilton, M. K., Ronveaux, C. C., Rust, B. M., Newman, J. W., Hawley, M., Barile, D., … Raybould, H. E. (2017). Prebiotic milk oligosaccharides prevent development of obese phenotype, impairment of gut permeability, and microbial dysbiosis in high fat-fed mice. American Journal of Physiology. Gastrointestinal and Liver Physiology, 312(5), G474–G487. https://doi.org/10.1152/ajpgi.00427.2016

He, H.-J., Qiao, J., Liu, Y., Guo, Q., Ou, X., & Wang, X. (2021). Isolation, structural, functional, and bioactive properties of cereal arabinoxylan—a critical review. Journal of Agricultural and Food Chemistry, 69(51), 15437 – 15457. https://doi.org/10.1021/acs.jafc.1c04506

Hernández-Ledesma, B., Hsieh, C.-C., & de Lumen, B. O. (2009a). Lunasin and Bowman-Birk protease inhibitor (BBI) in US commercial soy foods. Food chemistry, 115(2), 574–580. https://doi.org/10.1016/j.foodchem.2008.12.054

Hernández-Ledesma, B., Hsieh, C.-C., & de Lumen, B. O. (2009b). Lunasin, a novel seed peptide for cancer prevention. Peptides, 30(2), 426–430. https://doi.org/10.1016/j.peptides.2008.11.002

Hofman, D. L., van Buul, V. J., & Brouns, F. J. P. H. (2016). Nutrition, health, and regulatory aspects of digestible maltodextrins. Critical reviews in food science and nutrition, 56(12), 2091–2100. https://doi.org/10.1080/10408398.2014.940415

Holeček, M. (2018). Branched-chain amino acids in health and disease: metabolism, alterations in blood plasma, and as supplements. Nutrition & metabolism, 15, 33. https://doi.org/10.1186/s12986-018-0271-1

Holzhauser, T., Wackermann, O., Ballmer-Weber, B. K., Bindslev-Jensen, C., Scibilia, J., Perono-Garoffo, L., … Vieths, S. (2009). Soybean (Glycine max) allergy in Europe: Gly m 5 (beta-conglycinin) and Gly m 6 (glycinin) are potential diagnostic markers for severe allergic reactions to soy. The Journal of Allergy and Clinical Immunology, 123(2), 452–458. https://doi.org/10.1016/j.jaci.2008.09.034

Horiguchi, T., & Kitagishi, K. (1971). Studies on rice seed protease V. Protease inhibitor in rice seed1. Plant and Cell Physiology, 12(6), 907–915. https://doi.org/10.1093/oxfordjournals.pcp.a074695

Hou, A., Chen, P., Alloatti, J., Li, D., Mozzoni, L., Zhang, B., & Shi, A. (2009). Genetic variability of seed sugar content in worldwide soybean germplasm collections. Crop science, 49(3), 903–912. https://doi.org/10.2135/cropsci2008.05.0256

Hou, D. H., & Chang, S. K.-C. (2004). Structural characteristics of purified glycinin from soybeans stored under various conditions. Journal of Agricultural and Food Chemistry, 52(12), 3792–3800. https://doi.org/10.1021/jf035072z

House, J. D., Hill, K., Neufeld, J., Franczyk, A., & Nosworthy, M. G. (2019). Determination of the protein quality of almonds (Prunus dulcis L.) as assessed by in vitro and in vivo methodologies. Food science & nutrition, 7(9), 2932–2938. https://doi.org/10.1002/fsn3.1146

Hsieh, C.-C., Hernández-Ledesma, B., Jeong, H. J., Park, J. H., & de Lumen, B. O. (2010). Complementary roles in cancer prevention: protease inhibitor makes the cancer preventive peptide lunasin bioavailable. Plos One, 5(1), e8890. https://doi.org/10.1371/journal.pone.0008890

Hsieh, C.-C., Martínez-Villaluenga, C., de Lumen, B. O., & Hernández-Ledesma, B. (2018). Updating the research on the chemopreventive and therapeutic role of the peptide lunasin. Journal of the science of food and agriculture, 98(6), 2070–2079. https://doi.org/10.1002/jsfa.8719

Huang, Y.-P., & Lai, H.-M. (2016). Bioactive compounds and antioxidative activity of colored rice bran. Journal of food and drug analysis, 24(3), 564–574. https://doi.org/10.1016/j.jfda.2016.01.004

Huang, Y.-P., Paviani, B., Fukagawa, N. K., Phillips, K., & Barile, D. (2022). Comprehensive oligosaccharide profiling of commercial almond milk, soy milk, and soy flour. In preparation.

Hughes, S. A., Shewry, P. R., Li, L., Gibson, G. R., Sanz, M. L., & Rastall, R. A. (2007). In vitro fermentation by human fecal microflora of wheat arabinoxylans. Journal of Agricultural and Food Chemistry, 55(11), 4589–4595. https://doi.org/10.1021/jf070293g

Innova Market Insights. (2021). "Shared Planet" leads Innova Market Insights' Top Ten Trends for 2022. Retrieved June 30, 2022, from https://www.innovamarketinsights.com/press-release/shared-planet-leads-innova-market-insights-top-ten-trends-for-2022/

Islam, N., Shafiee, M., & Vatanparast, H. (2021). Trends in the consumption of conventional dairy milk and plant-based beverages and their contribution to nutrient intake among Canadians. Journal of human nutrition and dietetics : the official journal of the British Dietetic Association, 34(6), 1022–1034. https://doi.org/10.1111/jhn.12910

Ito, K., Sjölander, S., Sato, S., Movérare, R., Tanaka, A., Söderström, L., … Ebisawa, M. (2011). IgE to Gly m 5 and Gly m 6 is associated with severe allergic reactions to soybean in Japanese children. The Journal of Allergy and Clinical Immunology, 128(3), 673–675. https://doi.org/10.1016/j.jaci.2011.04.025

Jeong, H. J., Lam, Y., & de Lumen, B. O. (2002). Barley lunasin suppresses ras-induced colony formation and inhibits core histone acetylation in mammalian cells. Journal of Agricultural and Food Chemistry, 50(21), 5903–5908. https://doi.org/10.1021/jf0256945

Jeong, H. J., Jeong, J. B., Kim, D. S., Park, J. H., Lee, J. B., Kweon, D.-H., … de Lumen, B. O. (2007). The cancer preventive peptide lunasin from wheat inhibits core histone acetylation. Cancer Letters, 255(1), 42–48. https://doi.org/10.1016/j.canlet.2007.03.022

Jindal, S., Soni, G. L., & Singh, R. (1982). Effect of Feeding of Lectins from Lentils and Peas on the Intestinal and Hepatic Enzymes of Albino Rats. Journal of Plant Foods, 4(2), 95–103. https://doi.org/10.1080/0142968X.1982.11904253

Johansson, L., Virkki, L., Maunu, S., Lehto, M., Ekholm, P., & Varo, P. (2000). Structural characterization of water soluble β-glucan of oat bran. Carbohydrate polymers, 42(2), 143–148. https://doi.org/10.1016/S0144-8617(99)00157-5

Jones, D. A., DuPont, M. S., Ambrose, M. J., Frias, J., & Hedley, C. L. (1999). The discovery of compositional variation for the raffinose family of oligosaccharides in pea seeds. Seed science research, 9(4), 305–310. https://doi.org/10.1017/S0960258599000318

Jones, D. B. (1931). Factors for converting percentages of nitrogen in foods and feeds into percentages of proteins. No. 183. US Department of Agriculture.

Jones, J. M. (2014). CODEX-aligned dietary fiber definitions help to bridge the “fiber gap”. Nutrition Journal, 13, 34. https://doi.org/10.1186/1475-2891-13-34

Jood, S., Mehta, U., Singh, R., & Bhat, C. M. (1985). Effect of processing on flatus-producing factors in legumes. Journal of Agricultural and Food Chemistry, 33(2), 268–271. https://doi.org/10.1021/jf00062a028

Jordinson, M., Deprez, P. H., Playford, R. J., Heal, S., Freeman, T. C., Alison, M., & Calam, J. (1996). Soybean lectin stimulates pancreatic exocrine secretion via CCK-A receptors in rats. The American Journal of Physiology, 270(4 Pt 1), G653–9. https://doi.org/10.1152/ajpgi.1996.270.4.G653

Kabasser, S., Hafner, C., Chinthrajah, S., Sindher, S. B., Kumar, D., Kost, L. E., … Bublin, M. (2021). Identification of Pru du 6 as a potential marker allergen for almond allergy. Allergy, 76(5), 1463–1472. https://doi.org/10.1111/all.14613

Kabel, M. A., Kortenoeven, L., Schols, H. A., & Voragen, A. G. J. (2002). In vitro fermentability of differently substituted xylo-oligosaccharides. Journal of Agricultural and Food Chemistry, 50(21), 6205–6210. https://doi.org/10.1021/jf020220r

Karppinen, S., Kiiliäinen, K., Liukkonen, K., Forssell, P., & Poutanen, K. (2001). Extraction and in vitro Fermentation of Rye Bran Fractions. Journal of cereal science, 34(3), 269–278. https://doi.org/10.1006/jcrs.2001.0413

Kawakatsu, T., Hirose, S., Yasuda, H., & Takaiwa, F. (2010). Reducing rice seed storage protein accumulation leads to changes in nutrient quality and storage organelle formation. Plant Physiology, 154(4), 1842–1854. https://doi.org/10.1104/pp.110.164343

Kawakatsu, T., & Takaiwa, F. (2019). Rice proteins and essential amino acids. In Rice (pp. 109–130). Elsevier. https://doi.org/10.1016/B978-0-12-811508-4.00004-6

Kawamura, S., Nagao, K., & Kasai, T. (1977). Determination of free monosaccharides and detection of sugar alcohols in mature soybean seeds. Journal of nutritional science and vitaminology, 23(3), 249–255. https://doi.org/10.3177/jnsv.23.249

Kim, H. J., & White, P. J. (2013). Impact of the molecular weight, viscosity, and solubility of β-glucan on in vitro oat starch digestibility. Journal of Agricultural and Food Chemistry, 61(13), 3270–3277. https://doi.org/10.1021/jf305348j

Kim, M. Y., Lee, S. H., Jang, G. Y., Park, H. J., Li, M., Kim, S., … Jeong, H. S. (2015). Effects of high hydrostatic pressure treatment on the enhancement of functional components of germinated rough rice (Oryza sativa L.). Food chemistry, 166, 86–92. https://doi.org/10.1016/j.foodchem.2014.05.150

Kinsella, J. E. (1979). Functional properties of soy proteins. Journal of the American Oil Chemists' Society, 56(3Part1), 242–258. https://doi.org/10.1007/BF02671468

Klose, C., & Arendt, E. K. (2012). Proteins in oats; their synthesis and changes during germination: a review. Critical reviews in food science and nutrition, 52(7), 629–639. https://doi.org/10.1080/10408398.2010.504902

Koopman, R., Crombach, N., Gijsen, A. P., Walrand, S., Fauquant, J., Kies, A. K., … van Loon, L. J. C. (2009). Ingestion of a protein hydrolysate is accompanied by an accelerated in vivo digestion and absorption rate when compared with its intact protein. The American Journal of Clinical Nutrition, 90(1), 106–115. https://doi.org/10.3945/ajcn.2009.27474

Krishnan, H. B. (2000). Biochemistry and molecular biology of soybean seed storage proteins. Journal of New Seeds, 2(3), 1–25. https://doi.org/10.1300/J153v02n03_01

Kumar, V., Rani, A., Mittal, P., & Shuaib, M. (2019). Kunitz trypsin inhibitor in soybean: contribution to total trypsin inhibitor activity as a function of genotype and fate during processing. Journal of Food Measurement and Characterization, 13(2), 1583–1590. https://doi.org/10.1007/s11694-019-00074-y

Kuo, H.-Y., Chen, S.-H., & Yeh, A.-I. (2014). Preparation and physicochemical properties of whole-bean soymilk. Journal of Agricultural and Food Chemistry, 62(3), 742–749. https://doi.org/10.1021/jf404465w

Kuo, T. M., VanMiddlesworth, J. F., & Wolf, W. J. (1988). Content of raffinose oligosaccharides and sucrose in various plant seeds. Journal of Agricultural and Food Chemistry, 36(1), 32–36. https://doi.org/10.1021/jf00079a008

Kwak, J. H., Kim, M., Lee, E., Lee, S.-H., Ahn, C.-W., & Lee, J. H. (2013). Effects of black soy peptide supplementation on blood pressure and oxidative stress: a randomized controlled trial. Hypertension Research, 36(12), 1060–1066. https://doi.org/10.1038/hr.2013.79

Kwok, K. C., Qin, W. H., & Tsang, J. C. (1993). Heat Inactivation of Trypsin Inhibitors in Soymilk at Ultra-High Temperatures. Journal of Food Science, 58(4), 859–862. https://doi.org/10.1111/j.1365-2621.1993.tb09377.x

Lagos, L. V., & Stein, H. H. (2017). Chemical composition and amino acid digestibility of soybean meal produced in the United States, China, Argentina, Brazil, or India. Journal of Animal Science, 95(4), 1626. https://doi.org/10.2527/jas2017.1440

Lai, H.-M., & Huang, Y.-P. (2014). Effects of ball milling on the properties of colored rice bran. In P. Williams & G. Phillips (eds.), Gums and stabilisers for the food industry 17 (pp. 155–163). Cambridge: Royal Society of Chemistry. https://doi.org/10.1039/9781782621300-00155

Lajolo, F. M., & Genovese, M. I. (2002). Nutritional significance of lectins and enzyme inhibitors from legumes. Journal of Agricultural and Food Chemistry, 50(22), 6592–6598. https://doi.org/10.1021/jf020191k

Lam, A. C. Y., Can Karaca, A., Tyler, R. T., & Nickerson, M. T. (2016). Pea protein isolates: Structure, extraction, and functionality. Food Reviews International, 34(2), 1–22. https://doi.org/10.1080/87559129.2016.1242135

Lee, Y., Lee, H.-J., Lee, H.-S., Jang, Y.-A., & Kim, C. (2008). Analytical dietary fiber database for the National Health and Nutrition Survey in Korea. Journal of Food Composition and Analysis, 21, S35–S42. https://doi.org/10.1016/j.jfca.2007.07.008

Li, G.-H., Qu, M.-R., Wan, J.-Z., & You, J.-M. (2007). Antihypertensive effect of rice protein hydrolysate with in vitro angiotensin I-converting enzyme inhibitory activity in spontaneously hypertensive rats. Asia Pacific journal of clinical nutrition, 16 Suppl 1, 275–280.

Li, H., & Aluko, R. E. (2010). Identification and inhibitory properties of multifunctional peptides from pea protein hydrolysate. Journal of Agricultural and Food Chemistry, 58(21), 11471–11476. https://doi.org/10.1021/jf102538g

Li, H., Prairie, N., Udenigwe, C. C., Adebiyi, A. P., Tappia, P. S., Aukema, H. M., … Aluko, R. E. (2011). Blood pressure lowering effect of a pea protein hydrolysate in hypertensive rats and humans. Journal of Agricultural and Food Chemistry, 59(18), 9854–9860. https://doi.org/10.1021/jf201911p

Li, Q., Zhang, L., & Lametsch, R. (2020). Current progress in kokumi-active peptides, evaluation and preparation methods: a review. Critical reviews in food science and nutrition, 1–12. https://doi.org/10.1080/10408398.2020.1837726

Li, S., Bu, T., Zheng, J., Liu, L., He, G., & Wu, J. (2019). Preparation, Bioavailability, and Mechanism of Emerging Activities of Ile-Pro-Pro and Val-Pro-Pro. Comprehensive Reviews in Food Science and Food Safety, 18(4), 1097–1110. https://doi.org/10.1111/1541-4337.12457

Li, Y., Chen, M., Deng, L., Liang, Y., Liu, Y., Liu, W., … Liu, C. (2021). Whole soybean milk produced by a novel industry-scale micofluidizer system without soaking and filtering. Journal of food engineering, 291, 110228. https://doi.org/10.1016/j.jfoodeng.2020.110228

Lin, Y.-H., Li, H.-T., Huang, Y.-C., Hsieh, Y.-C., Guan, H.-H., Liu, M.-Y., … Chen, C.-J. (2006). Purification, crystallization and preliminary X-ray crystallographic analysis of rice Bowman-Birk inhibitor from Oryza sativa. Acta Crystallographica. Sect. F, Structural Biology and Crystallization Communications, 62(Pt 6), 522–524. https://doi.org/10.1107/S1744309106014795

Lis, H., & Sharon, N. (1973). The biochemistry of plant lectins (phytohemagglutinins). Annual Review of Biochemistry, 42(0), 541–574. https://doi.org/10.1146/annurev.bi.42.070173.002545

Liu, C., & Sathe, S. K. (2018). Food allergen epitope mapping. Journal of Agricultural and Food Chemistry, 66(28), 7238–7248. https://doi.org/10.1021/acs.jafc.8b01967

Liu, S., Proudman, J., & Mitloehner, F. M. (2021). Rethinking methane from animal agriculture. CABI Agriculture and Bioscience, 2(1), 22. https://doi.org/10.1186/s43170-021-00041-y

Lomer, M. C. E., Parkes, G. C., & Sanderson, J. D. (2008). Review article: lactose intolerance in clinical practice--myths and realities. Alimentary Pharmacology & Therapeutics, 27(2), 93–103. https://doi.org/10.1111/j.1365-2036.2007.03557.x

MacArthur, L. A., & D'appolonia, B. L. (1979). Comparison of oat and wheat carbohydrates. I. Sugars. Cereal Chem, 56(5), 455–457.

Maeda, H., & Nakamura, A. (2021). Soluble soybean polysaccharide. In Handbook of Hydrocolloids (pp. 463–480). Elsevier. https://doi.org/10.1016/B978-0-12-820104-6.00025-5

Mäkinen, O. E., Wanhalinna, V., Zannini, E., & Arendt, E. K. (2016). Foods for Special Dietary Needs: Non-dairy Plant-based Milk Substitutes and Fermented Dairy-type Products. Critical reviews in food science and nutrition, 56(3), 339–349. https://doi.org/10.1080/10408398.2012.761950

Manthey, F. A., Hareland, G. A., & Huseby, D. J. (1999). Soluble and insoluble dietary fiber content and composition in oat. Cereal Chemistry Journal, 76(3), 417–420. https://doi.org/10.1094/CCHEM.1999.76.3.417

Marinangeli, C. P. F., & House, J. D. (2017). Potential impact of the digestible indispensable amino acid score as a measure of protein quality on dietary regulations and health. Nutrition Reviews, 75(8), 658–667. https://doi.org/10.1093/nutrit/nux025

Mariotti, F., Tomé, D., & Mirand, P. P. (2008). Converting nitrogen into protein-beyond 6.25 and Jones' factors. Critical reviews in food science and nutrition, 48(2), 177–184. https://doi.org/10.1080/10408390701279749

Martín-Cabrejas, M. A., Ariza, N., Esteban, R., Mollá, E., Waldron, K., & López-Andréu, F. J. (2003). Effect of germination on the carbohydrate composition of the dietary fiber of peas (Pisum sativum L.). Journal of Agricultural and Food Chemistry, 51(5), 1254–1259. https://doi.org/10.1021/jf0207631

Maruyama, N., Katsube, T., Wada, Y., Oh, M. H., Barba De La Rosa, A. P., Okuda, E., … Utsumi, S. (1998). The roles of the N-linked glycans and extension regions of soybean beta-conglycinin in folding, assembly and structural features. European Journal of Biochemistry / FEBS, 258(2), 854–862. https://doi.org/10.1046/j.1432-1327.1998.2580854.x

Mathai, J. K., Liu, Y., & Stein, H. H. (2017). Values for digestible indispensable amino acid scores (DIAAS) for some dairy and plant proteins may better describe protein quality than values calculated using the concept for protein digestibility-corrected amino acid scores (PDCAAS). The British Journal of Nutrition, 117(4), 490–499. https://doi.org/10.1017/S0007114517000125

McCarthy, K. S., Parker, M., Ameerally, A., Drake, S. L., & Drake, M. A. (2017). Drivers of choice for fluid milk versus plant-based alternatives: What are consumer perceptions of fluid milk? Journal of Dairy Science, 100(8), 6125–6138. https://doi.org/10.3168/jds.2016-12519

McJarrow, P., & van Amelsfort-Schoonbeek, J. (2004). Bovine sialyl oligosaccharides: seasonal variations in their concentrations in milk, and a comparison of the colostrums of Jersey and Friesian cows. International Dairy Journal, 14(7), 571–579. https://doi.org/10.1016/j.idairyj.2003.11.006

Messina, M., & Venter, C. (2020). Recent surveys on food allergy prevalence. Nutrition today, 55(1), 22–29. https://doi.org/10.1097/NT.0000000000000389

Millward, D. J. (1997). Human amino acid requirements. The Journal of Nutrition, 127(9), 1842–1846. https://doi.org/10.1093/jn/127.9.1842

Min, B., Kyung Koo, O., Park, S. H., Jarvis, N., Ricke, S. C., Crandall, P. G., & Lee, S.-O. (2015). Fermentation patterns of various pectin sources by human fecal microbiota. Food and Nutrition Sciences, 06(12), 1103–1114. https://doi.org/10.4236/fns.2015.612115

Minkiewicz, P., Iwaniak, A., & Darewicz, M. (2019). BIOPEP-UWM Database of Bioactive Peptides: Current Opportunities. International Journal of Molecular Sciences, 20(23). https://doi.org/10.3390/ijms20235978

Mishkind, M. L., Palevitz, B. A., Raikhel, N. V., & Keegstra, K. (1983). Localization of wheat germ agglutinin--like lectins in various species of the gramineae. Science, 220(4603), 1290–1292. https://doi.org/10.1126/science.220.4603.1290

Mitchell, R. A. C., Lovegrove, A., & Shewry, P. R. (2013). Lunasin in cereal seeds: What is the origin? Journal of cereal science, 57(3), 267–269. https://doi.org/10.1016/j.jcs.2013.01.013

Miyoshi, N., Koyama, Y., Katsuno, Y., Hayakawa, S., Mita, T., Ohta, T., … Isemura, M. (2001). Apoptosis induction associated with cell cycle dysregulation by rice bran agglutinin. Journal of Biochemistry, 130(6), 799–805. https://doi.org/10.1093/oxfordjournals.jbchem.a003051

Mohan, B. H., Malleshi, N. G., & Koseki, T. (2010). Physico-chemical characteristics and non-starch polysaccharide contents of Indica and Japonica brown rice and their malts. LWT - Food Science and Technology, 43(5), 784–791. https://doi.org/10.1016/j.lwt.2010.01.002

Mohnen, D. (2008). Pectin structure and biosynthesis. Current Opinion in Plant Biology, 11(3), 266–277. https://doi.org/10.1016/j.pbi.2008.03.006

Muench, D. G., & Okita, T. W. (1997). The storage proteins of rice and oat. In B. A. Larkins & I. K. Vasil (eds.), Cellular and molecular biology of plant seed development (Vol. 4, pp. 289–330). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-015-8909-3_8

Muramoto, K. (2017). Lectins as bioactive proteins in foods and feeds. Food science and technology research, 23(4), 487–494. https://doi.org/10.3136/fstr.23.487

Nakamura, Y., Yamamoto, N., Sakai, K., Okubo, A., Yamazaki, S., & Takano, T. (1995). Purification and characterization of angiotensin I-converting enzyme inhibitors from sour milk. Journal of Dairy Science, 78(4), 777–783. https://doi.org/10.3168/jds.S0022-0302(95)76689-9

Nakamura, Y., Yamamoto, N., Sakai, K., & Takano, T. (1995). Antihypertensive effect of sour milk and peptides isolated from it that are inhibitors to angiotensin I-converting enzyme. Journal of Dairy Science, 78(6), 1253–1257. https://doi.org/10.3168/jds.S0022-0302(95)76745-5

Nakata, H., Lin, C. Y., Abolhassani, M., Ogawa, T., Tateno, H., Hirabayashi, J., & Muramoto, K. (2017). Isolation of Rice Bran Lectins and Characterization of Their Unique Behavior in Caco-2 Cells. International Journal of Molecular Sciences, 18(5). https://doi.org/10.3390/ijms18051052

Nakurte, I., Kirhnere, I., Namniece, J., Saleniece, K., Krigere, L., Mekss, P., … Muceniece, R. (2013). Detection of the lunasin peptide in oats (Avena sativa L). Journal of cereal science, 57(3), 319–324. https://doi.org/10.1016/j.jcs.2012.12.008

Nakurte, I., Klavins, K., Kirhnere, I., Namniece, J., Adlere, L., Matvejevs, J., … Muceniece, R. (2012). Discovery of lunasin peptide in triticale (X Triticosecale Wittmack). Journal of cereal science, 56(2), 510–514. https://doi.org/10.1016/j.jcs.2012.04.004

Neyrinck, A. M., Possemiers, S., Druart, C., Van deaaa 1Wiele, T., De Backer, F., Cani, P. D., … Delzenne, N. M. (2011). Prebiotic effects of wheat arabinoxylan related to the increase in bifidobacteria, Roseburia and Bacteroides/Prevotella in diet-induced obese mice. Plos One, 6(6), e20944. https://doi.org/10.1371/journal.pone.0020944

Odani, S., & Ikenaka, T. (1977). Studies on soybean trypsin inhibitors. XI. Complete amino acid sequence of a soybean trypsin-chymotrypsin-elastase inhibitor, C-II. Journal of Biochemistry, 82(6), 1523–1531. https://doi.org/10.1093/oxfordjournals.jbchem.a131846

Ohlsson, J. A., Johansson, M., Hansson, H., Abrahamson, A., Byberg, L., Smedman, A., … Lundh, Å. (2017). Lactose, glucose and galactose content in milk, fermented milk and lactose-free milk products. International Dairy Journal, 73, 151–154. https://doi.org/10.1016/j.idairyj.2017.06.004

Park, J. H., Jeong, H. J., & Lumen, B. O. de. (2007). In vitro digestibility of the cancer-preventive soy peptides lunasin and BBI. Journal of Agricultural and Food Chemistry, 55(26), 10703–10706. https://doi.org/10.1021/jf072107c

Pastell, H., Virkki, L., Harju, E., Tuomainen, P., & Tenkanen, M. (2009). Presence of 1-->3-linked 2-O-beta-d-xylopyranosyl-alpha-l-arabinofuranosyl side chains in cereal arabinoxylans. Carbohydrate Research, 344(18), 2480–2488. https://doi.org/10.1016/j.carres.2009.09.035

Peterson, D M. (1978). Subunit structure and composition of oat seed globulin. Plant Physiology, 62(4), 506–509. https://doi.org/10.1104/pp.62.4.506

Peterson, David M. (1992). Composition and nutritional characteristics of oat grain and products. In H. G. Marshall & M. E. Sorrells (eds.), Oat science and technology (pp. 265–292). Madison, WI, USA: American Society of Agronomy, Crop Science Society of America. https://doi.org/10.2134/agronmonogr33.c10

Pettersson, Å., Lindberg, J. E., Thomke, S., & Eggum, B. O. (1996). Nutrient digestibility and protein quality of oats differing in chemical composition evaluated in rats and by an in vitro technique. Animal feed science and technology, 62(2-4), 203–213. https://doi.org/10.1016/S0377-8401(96)00964-9

Pew Research Center. (2021, May 26). Gen Z, Millennials Stand Out for Climate Change Activism, Social Media Engagement With Issue | . https://www.pewresearch.org/science/2021/05/26/gen-z-millennials-stand-out-for-climate-change-activism-social-media-engagement-with-issue/

Phuwadolpaisarn, P. (2021). Comparison of β-Glucan Content in Milled Rice, Rice Husk and Rice Bran from Rice Cultivars Grown in Different Locations of Thailand and the Relationship between β-Glucan and Amylose Contents. Molecules (Basel, Switzerland), 26(21). https://doi.org/10.3390/molecules26216368

Pi, X., Sun, Y., Fu, G., Wu, Z., & Cheng, J. (2021). Effect of processing on soybean allergens and their allergenicity. Trends in food science & technology, 118, 316–327. https://doi.org/10.1016/j.tifs.2021.10.006

Picolli da Silva, L., & de Lourdes Santorio Ciocca, M. (2005). Total, insoluble and soluble dietary fiber values measured by enzymatic–gravimetric method in cereal grains. Journal of Food Composition and Analysis, 18(1), 113–120. https://doi.org/10.1016/j.jfca.2003.12.005

Písaříková, B., & Zralý, Z. (2010). Dietary Fibre Content in Lupine (Lupinus albus L.) and Soya (Glycine max L.) Seeds. Acta Veterinaria Brno, 79(2), 211–216. https://doi.org/10.2754/avb201079020211

Poola, I. (1989). Rice lectin: Physico-chemical and carbohydrate-binding properties. Carbohydrate polymers, 10(4), 281–288. https://doi.org/10.1016/0144-8617(89)90067-2

Prakash, B., Selvaraj, S., Murthy, M. R., Sreerama, Y. N., Rao, D. R., & Gowda, L. R. (1996). Analysis of the amino acid sequences of plant Bowman-Birk inhibitors. Journal of Molecular Evolution, 42(5), 560–569. https://doi.org/10.1007/BF02352286

Pusztai, A., Ewen, S. W., Grant, G., Brown, D. S., Stewart, J. C., Peumans, W. J., … Bardocz, S. (1993). Antinutritive effects of wheat-germ agglutinin and other N-acetylglucosamine-specific lectins. The British Journal of Nutrition, 70(1), 313–321. https://doi.org/10.1079/bjn19930124

Pusztai, A., Ewen, S. W., Grant, G., Peumans, W. J., van Damme, E. J., Rubio, L., & Bardocz, S. (1990). Relationship between survival and binding of plant lectins during small intestinal passage and their effectiveness as growth factors. Digestion, 46 Suppl 2, 308–316. https://doi.org/10.1159/000200402

Quillien, L., Ferrasson, E., Molle, D., & Gueguen, J. (1997). Trypsin inhibitor polymorphism: multigene family expression and posttranslational modification. Journal of protein chemistry, 16(3), 195–203. https://doi.org/10.1023/a:1026326808553

Rackis, J. J., Wolf, W. J., & Baker, E. C. (1986). Protease inhibitors in plant foods: content and inactivation. Advances in Experimental Medicine and Biology, 199, 299–347. https://doi.org/10.1007/978-1-4757-0022-0_19

Radauer, C., Nandy, A., Ferreira, F., Goodman, R. E., Larsen, J. N., Lidholm, J., … Breiteneder, H. (2014). Update of the WHO/IUIS Allergen Nomenclature Database based on analysis of allergen sequences. Allergy, 69(4), 413–419. https://doi.org/10.1111/all.12348

Ramulu, P., & Rao, P. U. (1997). Effect of processing on dietary fiber content of cereals and pulses. Plant foods for human nutrition (Dordrecht, Netherlands), 50(3), 249–257. https://doi.org/10.1007/BF02436061

Real, A., Comino, I., de Lorenzo, L., Merchán, F., Gil-Humanes, J., Giménez, M. J., … Pistón, F. (2012). Molecular and immunological characterization of gluten proteins isolated from oat cultivars that differ in toxicity for celiac disease. Plos One, 7(12), e48365. https://doi.org/10.1371/journal.pone.0048365

Reynaud, Y., Buffière, C., Cohade, B., Vauris, M., Liebermann, K., Hafnaoui, N., … Rémond, D. (2021). True ileal amino acid digestibility and digestible indispensable amino acid scores (DIAASs) of plant-based protein foods. Food chemistry, 338, 128020. https://doi.org/10.1016/j.foodchem.2020.128020

Rizzi, C., Galeoto, L., Zoccatelli, G., Vincenzi, S., Chignola, R., & Peruffo, A. D. . (2003). Active soybean lectin in foods: quantitative determination by ELISA using immobilised asialofetuin. Food Research International, 36(8), 815–821. https://doi.org/10.1016/S0963-9969(03)00076-0

Robinson, R. C. (2019). Structures and metabolic properties of bovine milk oligosaccharides and their potential in the development of novel therapeutics. Frontiers in nutrition, 6, 50. https://doi.org/10.3389/fnut.2019.00050

Rose, D. J., & Inglett, G. E. (2010). Production of feruloylated arabinoxylo-oligosaccharides from maize (Zea mays) bran by microwave-assisted autohydrolysis. Food chemistry, 119(4), 1613–1618. https://doi.org/10.1016/j.foodchem.2009.09.053

Rutherfurd, S. M., Fanning, A. C., Miller, B. J., & Moughan, P. J. (2015). Protein digestibility-corrected amino acid scores and digestible indispensable amino acid scores differentially describe protein quality in growing male rats. The Journal of Nutrition, 145(2), 372–379. https://doi.org/10.3945/jn.114.195438

Sadeghalvad, M., Mohammadi-Motlagh, H.-R., Karaji, A. G., & Mostafaie, A. (2019). In vivo anti-inflammatory efficacy of the combined Bowman-Birk trypsin inhibitor and genistein isoflavone, two biological compounds from soybean. Journal of Biochemical and Molecular Toxicology, 33(12), e22406. https://doi.org/10.1002/jbt.22406

Safavi, F., & Rostami, A. (2012). Role of serine proteases in inflammation: Bowman-Birk protease inhibitor (BBI) as a potential therapy for autoimmune diseases. Experimental and Molecular Pathology, 93(3), 428–433. https://doi.org/10.1016/j.yexmp.2012.09.014

Sajadimajd, S., Bahrami, G., Daglia, M., Nabavi, S. M., Naseri, R., & Farzaei, M. H. (2019). Plant-Derived Supplementary Carbohydrates, Polysaccharides and Oligosaccharides in Management of Diabetes Mellitus: A Comprehensive Review. Food Reviews International, 35(6), 563–586. https://doi.org/10.1080/87559129.2019.1584818

Sánchez, A., & Vázquez, A. (2017). Bioactive peptides: A review. Food Quality and Safety, 1(1), 29–46. https://doi.org/10.1093/fqsafe/fyx006

Sathe, S. K., Wolf, W. J., Roux, K. H., Teuber, S. S., Venkatachalam, M., & Sze-Tao, K. W. C. (2002). Biochemical characterization of amandin, the major storage protein in almond (Prunus dulcis L.). Journal of Agricultural and Food Chemistry, 50(15), 4333–4341. https://doi.org/10.1021/jf020007v

Sato, K. (2018). Structure, Content, and Bioactivity of Food-Derived Peptides in the Body. Journal of Agricultural and Food Chemistry, 66(12), 3082–3085. https://doi.org/10.1021/acs.jafc.8b00390

Sethi, S., Tyagi, S. K., & Anurag, R. K. (2016). Plant-based milk alternatives an emerging segment of functional beverages: a review. Journal of food science and technology, 53(9), 3408–3423. https://doi.org/10.1007/s13197-016-2328-3

Shen, W., & Matsui, T. (2017). Current knowledge of intestinal absorption of bioactive peptides. Food & function, 8(12), 4306–4314. https://doi.org/10.1039/c7fo01185g

Shewry, P. R., Napier, J. A., & 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., & Halford, N. G. (2002). Cereal seed storage proteins: structures, properties and role in grain utilization. Journal of Experimental Botany, 53(370), 947–958. https://doi.org/10.1093/jexbot/53.370.947

Shibata, M., Hirotsuka, M., Mizutani, Y., Takahashi, H., Kawada, T., Matsumiya, K., … Matsumura, Y. (2017). Isolation and characterization of key contributors to the “kokumi” taste in soybean seeds. Bioscience, Biotechnology, and Biochemistry, 81(11), 2168–2177. https://doi.org/10.1080/09168451.2017.1372179

Shibuya, N. (1984). Phenolic acids and their carbohydrate esters in rice endosperm cell walls. Phytochemistry, 23(10), 2233–2237. https://doi.org/10.1016/S0031-9422(00)80526-9

Shibuya, N. (1989). Comparative Studies on the Cell Wall Polymers Obtained from Different Parts of Rice Grains. In N. G. Lewis & M. G. Paice (eds.), Plant cell wall polymers: biogenesis and biodegradation (Vol. 399, pp. 333–344). Washington, DC: American Chemical Society. https://doi.org/10.1021/bk-1989-0399.ch024

Silva, A. R. A., Silva, M. M. N., & Ribeiro, B. D. (2020). Health issues and technological aspects of plant-based alternative milk. Food research international (Ottawa, Ont.), 131, 108972. https://doi.org/10.1016/j.foodres.2019.108972

Singh Sindhu, A., Zheng, Z., & Murai, N. (1997). The pea seed storage protein legumin was synthesized, processed, and accumulated stably in transgenic rice endosperm. Plant Science, 130(2), 189–196. https://doi.org/10.1016/S0168-9452(97)00219-7

Souza, T. S. P., Dias, F. F. G., Koblitz, M. G. B., & 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

Stone, A. K., Avarmenko, N. A., Warkentin, T. D., & Nickerson, M. T. (2015). Functional properties of protein isolates from different pea cultivars. Food science and biotechnology, 24(3), 827–833. https://doi.org/10.1007/s10068-015-0107-y

Tao, K., Yu, W., Prakash, S., & Gilbert, R. G. (2019). High-amylose rice: Starch molecular structural features controlling cooked rice texture and preference. Carbohydrate polymers, 219, 251–260. https://doi.org/10.1016/j.carbpol.2019.05.031

Tian, L., Gruppen, H., & Schols, H. A. (2015). Characterization of (Glucurono)arabinoxylans from Oats Using Enzymatic Fingerprinting. Journal of Agricultural and Food Chemistry, 63(50), 10822–10830. https://doi.org/10.1021/acs.jafc.5b04419

Tian, L., Scholte, J., Borewicz, K., van den Bogert, B., Smidt, H., Scheurink, A. J. W., … Schols, H. A. (2016). Effects of pectin supplementation on the fermentation patterns of different structural carbohydrates in rats. Molecular Nutrition & Food Research, 60(10), 2256–2266. https://doi.org/10.1002/mnfr.201600149

Tomatsu, M., Shimakage, A., Shinbo, M., Yamada, S., & Takahashi, S. (2013). Novel angiotensin I-converting enzyme inhibitory peptides derived from soya milk. Food chemistry, 136(2), 612–616. https://doi.org/10.1016/j.foodchem.2012.08.080

Tömösközi, S., Lásztity, R., Haraszi, R., & Baticz, O. (2001). Isolation and study of the functional properties of pea proteins. Food / Nahrung.

Tran, T. U., Suzuki, K., Okadome, H., Ikezaki, H., Homma, S., & Ohtsubo, K. (2005). Detection of changes in taste of japonica and indica brown and milled rice (Oryza sativa L.) during storage using physicochemical analyses and a taste sensing system. Journal of Agricultural and Food Chemistry, 53(4), 1108–1118. https://doi.org/10.1021/jf049064+

Trcka, J., Schäd, S. G., Scheurer, S., Conti, A., Vieths, S., Gross, G., & Trautmann, A. (2012). Rice-induced anaphylaxis: IgE-mediated allergy against a 56-kDa glycoprotein. International Archives of Allergy and Immunology, 158(1), 9–17. https://doi.org/10.1159/000330641

Truong, K. T. P., & Rumpagaporn, P. (2019). Oligosaccharides Preparation from Rice Bran Arabinoxylan by Two Different Commercial Endoxylanase Enzymes. Journal of nutritional science and vitaminology, 65(Supplement), S171–S174. https://doi.org/10.3177/jnsv.65.S171

U.S. Food and Drug Administration Center for Food Safety and Applied Nutrition. (2019). Guidance for Industry: Nutrition and Supplement Facts Labels Questions and Answers Related to the Compliance Date, Added Sugars, and Declaration of Quantitative Amounts of Vitamins and Minerals. U.S. Food and Drug Administration.

Udenigwe, C. C., Abioye, R. O., Okagu, I. U., & Obeme-Nmom, J. I. (2021). Bioaccessibility of bioactive peptides: recent advances and perspectives. Current Opinion in Food Science, 39, 182–189. https://doi.org/10.1016/j.cofs.2021.03.005

Undhad Trupti, J., Das, S., Solanki, D., Kinariwala, D., & Hati, S. (2021). Bioactivities and ACE-inhibitory peptides releasing potential of lactic acid bacteria in fermented soy milk. Food Production, Processing and Nutrition, 3(1), 10. https://doi.org/10.1186/s43014-021-00056-y

Urbano, G., López-Jurado, M., Frejnagel, S., Gómez-Villalva, E., Porres, J. M., Frías, J., … Aranda, P. (2005). Nutritional assessment of raw and germinated pea (Pisum sativum L.) protein and carbohydrate by in vitro and in vivo techniques. Nutrition, 21(2), 230–239. https://doi.org/10.1016/j.nut.2004.04.025

Valenta, R., Hochwallner, H., Linhart, B., & Pahr, S. (2015). Food allergies: the basics. Gastroenterology, 148(6), 1120–31.e4. https://doi.org/10.1053/j.gastro.2015.02.006

Van Craeyveld, V., Swennen, K., Dornez, E., Van deaaa 1Wiele, T., Marzorati, M., Verstraete, W., … Courtin, C. M. (2008). Structurally different wheat-derived arabinoxylooligosaccharides have different prebiotic and fermentation properties in rats. The Journal of Nutrition, 138(12), 2348–2355. https://doi.org/10.3945/jn.108.094367

Van den Abbeele, P., Gérard, P., Rabot, S., Bruneau, A., El Aidy, S., Derrien, M., … Possemiers, S. (2011). Arabinoxylans and inulin differentially modulate the mucosal and luminal gut microbiota and mucin-degradation in humanized rats. Environmental Microbiology, 13(10), 2667–2680. https://doi.org/10.1111/j.1462-2920.2011.02533.x

Vasconcelos, I. M., & Oliveira, J. T. A. (2004). Antinutritional properties of plant lectins. Toxicon, 44(4), 385–403. https://doi.org/10.1016/j.toxicon.2004.05.005

Wakasa, Y., Yang, L., Hirose, S., & Takaiwa, F. (2009). Expression of unprocessed glutelin precursor alters polymerization without affecting trafficking and accumulation. Journal of Experimental Botany, 60(12), 3503–3511. https://doi.org/10.1093/jxb/erp187

Walstra, P., Walstra, P., Wouters, J. T. M., & Geurts, T. J. (2005). Dairy science and technology. CRC Press. https://doi.org/10.1201/9781420028010

Wang, J., Bai, J., Fan, M., Li, T., Li, Y., Qian, H., … Rao, Z. (2020). Cereal-derived arabinoxylans: Structural features and structure–activity correlations. Trends in food science & technology, 96, 157–165. https://doi.org/10.1016/j.tifs.2019.12.016

Wang, M., Monaco, M. H., Hauser, J., Yan, J., Dilger, R. N., & Donovan, S. M. (2021). Bovine milk oligosaccharides and human milk oligosaccharides modulate the gut microbiota composition and volatile fatty acid concentrations in a preclinical neonatal model. Microorganisms, 9(5). https://doi.org/10.3390/microorganisms9050884

Wang, Q, Wood, P. J., Huang, X., & Cui, W. (2003). Preparation and characterization of molecular weight standards of low polydispersity from oat and barley (1→3)(1→4)-β-d-glucan. Food hydrocolloids, 17(6), 845 – 853. https://doi.org/10.1016/S0268-005X(03)00105-X

Wang, Qi, & Ellis, P. R. (2014). Oat β-glucan: physico-chemical characteristics in relation to its blood-glucose and cholesterol-lowering properties. The British Journal of Nutrition, 112 Suppl 2, S4–S13. https://doi.org/10.1017/S0007114514002256

Wen, T. N., & Luthe, D. S. (1985). Biochemical characterization of rice glutelin. Plant Physiology, 78(1), 172–177. https://doi.org/10.1104/pp.78.1.172

Westerlund, E., Andersson, R., & Åman, P. (1993). Isolation and chemical characterization of water-soluble mixed-linked β-glucans and arabinoxylans in oat milling fractions. Carbohydrate polymers, 20(2), 115–123. https://doi.org/10.1016/0144-8617(93)90086-J

Whitehead, A., Beck, E. J., Tosh, S., & Wolever, T. M. S. (2014). Cholesterol-lowering effects of oat β-glucan: a meta-analysis of randomized controlled trials. The American Journal of Clinical Nutrition, 100(6), 1413–1421. https://doi.org/10.3945/ajcn.114.086108

Wilson, L. A., Birmingham, V. A., Moon, D. P., & Snyder, H. E. (1978). Isolation and characterization of starch from mature soybeans. Cereal chemistry.

Wolever, T. M. S., Tosh, S. M., Gibbs, A. L., Brand-Miller, J., Duncan, A. M., Hart, V., … Wood, P. J. (2010). Physicochemical properties of oat β-glucan influence its ability to reduce serum LDL cholesterol in humans: a randomized clinical trial. The American Journal of Clinical Nutrition, 92(4), 723–732. https://doi.org/10.3945/ajcn.2010.29174

Wolf, C. A., Malone, T., & McFadden, B. R. (2020). Beverage milk consumption patterns in the United States: Who is substituting from dairy to plant-based beverages? Journal of Dairy Science, 103(12), 11209–11217. https://doi.org/10.3168/jds.2020-18741

Wu, G. (2010). Functional amino acids in growth, reproduction, and health. Advances in nutrition (Bethesda, Md.), 1(1), 31–37. https://doi.org/10.3945/an.110.1008

Wu, J., & Ding, X. (2001). Hypotensive and physiological effect of angiotensin converting enzyme inhibitory peptides derived from soy protein on spontaneously hypertensive rats. Journal of Agricultural and Food Chemistry, 49(1), 501–506. https://doi.org/10.1021/jf000695n

Yada, S., Huang, G., & Lapsley, K. (2013). Natural variability in the nutrient composition of California-grown almonds. Journal of Food Composition and Analysis, 30(2), 80–85. https://doi.org/10.1016/j.jfca.2013.01.008

Yamamoto, S., Tomiyama, M., Nemoto, R., Naganuma, T., Ogawa, T., & Muramoto, K. (2013). Effects of food lectins on the transport system of human intestinal Caco-2 cell monolayers. Bioscience, Biotechnology, and Biochemistry, 77(9), 1917–1924. https://doi.org/10.1271/bbb.130367

Yang, H.-Y., Yang, S.-C., Chen, S.-T., & Chen, J.-R. (2008). Soy protein hydrolysate ameliorates cardiovascular remodeling in rats with L-NAME-induced hypertension. The Journal of Nutritional Biochemistry, 19(12), 833–839. https://doi.org/10.1016/j.jnutbio.2007.11.004

Yang, J., Zamani, S., Liang, L., & Chen, L. (2021). Extraction methods significantly impact pea protein composition, structure and gelling properties. Food hydrocolloids, 117, 106678. https://doi.org/10.1016/j.foodhyd.2021.106678

Yang, J., Bai, W., Zeng, X., & Cui, C. (2019). Gamma glutamyl peptides: The food source, enzymatic synthesis, kokumi-active and the potential functional properties – A review. Trends in food science & technology, 91, 339–346. https://doi.org/10.1016/j.tifs.2019.07.022

Yuan, S., Chang, S. K. C., Liu, Z., & Xu, B. (2008). Elimination of trypsin inhibitor activity and beany flavor in soy milk by consecutive blanching and ultrahigh-temperature (UHT) processing. Journal of Agricultural and Food Chemistry, 56(17), 7957–7963. https://doi.org/10.1021/jf801039h

Zhang, K., Dong, R., Hu, X., Ren, C., & Li, Y. (2021). Oat-Based Foods: Chemical Constituents, Glycemic Index, and the Effect of Processing. Foods, 10(6). https://doi.org/10.3390/foods10061304

Zhou, X.-L., Kong, X.-F., Lian, G.-Q., Blachier, F., Geng, M.-M., & Yin, Y.-L. (2014). Dietary supplementation with soybean oligosaccharides increases short-chain fatty acids but decreases protein-derived catabolites in the intestinal luminal content of weaned Huanjiang mini-piglets. Nutrition research (New York, N.Y.), 34(9), 780–788. https://doi.org/10.1016/j.nutres.2014.08.008

Zhou, Z., Robards, K., Helliwell, S., & Blanchard, C. (2002). Composition and functional properties of rice. International Journal of Food Science and Technology, 37(8), 849–868. https://doi.org/10.1046/j.1365-2621.2002.00625.x

Zurbau, A., Noronha, J. C., Khan, T. A., Sievenpiper, J. L., & Wolever, T. M. S. (2021). The effect of oat β-glucan on postprandial blood glucose and insulin responses: a systematic review and meta-analysis. European Journal of Clinical Nutrition, 75(11), 1540–1554. https://doi.org/10.1038/s41430-021-00875-9

Chapter II

Food glycomics: Dealing with unexpected degradation of oligosaccharides during sample preparation and analysis

(This chapter was published as a journal article "**Huang, Y.-P**.; Robinson, R.C.; Barile, D. *J. Food Drug Anal.* 2022, 30, 62–76, doi:10.38212/2224-6614.3393.")

Abstract

This study reveals that unexpected degradation of food oligosaccharides can occur during conventional glycomics workflows, including sample preparation and analysis by liquid chromatography-mass spectrometry (LC-MS). With the present investigation, we aim to alert the scientific community of the susceptibility of specific glycosidic linkages to degradation induced by heat and acid. Key standard oligosaccharides representing the major types found in foods (3'-sialyllactose and 6'-sialyl-N-acetyllactosamine for milk, raffinose and stachyose for legumes) were selected as model systems and underwent each of the following treatments independently: (1) labeled with the derivatizing agent 1-aminopyrene-3,6,8-trisulfonic (APTS) (followed by analysis with a capillary electrophoresis system coupled with a fluorescence detector), (2) dried from an acetonitrile-water mixture containing 0.1% trifluoroacetic acid, and (3) injected into an LC-MS system. We demonstrated that both raffinose and stachyose degraded during APTS-labeling by the acid in the labeling reagents. We also discovered that during centrifugal evaporation at 37°C, all of the four nonderivatized oligosaccharides tested were partially degraded. Additionally, when the LC-MS eluent contained 0.1% formic acid, 3'-sialyllactose, raffinose, and stachyose underwent extensive in-source fragmentation during analysis. Lastly, we identified a simple strategy that can reduce the probability of incorrect oligosaccharide identification resulting from extensive in-source fragmentation.

Keywords: In-source fragmentation · Oligosaccharide degradation · Raffinose-family oligosaccharides · Reductive amination · Sialylated oligosaccharides

2.1. Introduction

Oligosaccharides (OS) are carbohydrates composed of 3–20 monosaccharide moieties. They naturally exist in free form in foods, such as milk, legumes, honey, and vegetables, and can be generated through processing techniques, including enzymatic synthesis, depolymerization of polysaccharides, and enzymatic release from glycoproteins [1,2]. Glycosyltransferases are commonly adopted for OS synthesis using simple sugars as substrates, such as in the production of galactooligosaccharides and fructooligosaccharides [3]. Polysaccharide depolymerization can be fulfilled enzymatically, chemically, or physically, as seen in the production of fructooligosaccharides from inulin and xylooligosaccharides from xylan [4–7]. N-Glycosylation is a co/post-translational modification. Glycoproteins with N-glycosylation can release N-glycans via enzymatic treatment using specific enzymes and have become an emerging source of OS [2,8].

Some OS that are resistant to digestion can be utilized by commensal bacteria in the human gut and selectively stimulate the growth of beneficial intestinal bacteria, including species from the genera *Bifidobacterium*, *Lactobacillus*, and *Eubacterium*. In turn, they can render health benefits to the host, such as increasing mineral absorption, regulating blood lipid and blood glucose, reducing the risk of colon cancer, and modulating immune function [9]. The fermentability of different types of OS that originate from various sources or are produced via different processing techniques may vary and depends on whether the intestinal bacteria can generate the corresponding glycolytic enzymes to cleave the specific glycosidic linkages and whether the bacteria can further utilize the released monosaccharides. Therefore, collecting detailed information on OS structures, including monosaccharide units, degree of polymerization (DP), and types of glycosidic linkages, is critical when studying and predicting their fermentability, prebiotic properties, and structure-function relationships.

Acids are frequently used in the processing and analysis of food carbohydrates. Several of these techniques are based on the hydrolysis of glycosidic linkages in carbohydrates. For example, hydrolysis using hydrochloric or sulfuric acid at elevated temperatures (above 60 °C) is used for producing oligosaccharides with potential prebiotic property from polysaccharides on a large scale [10]. Dilute hydrochloric or sulfuric acid is also used for treating starch granules to modify starch structure and functionality [11]. Acid hydrolysis is also a crucial step in monosaccharide composition analysis for breaking down carbohydrates with various DP into constituent monosaccharides using sulfuric or trifluoroacetic acid (TFA) [12,13]. Dilute acids can serve other purposes in the analysis of food carbohydrates that are unrelated to OS hydrolysis, such as functioning as an electronic modifier or pH modifier in solid-phase extraction and liquid chromatography [14]. Furthermore, as carbohydrates are invisible to UV-spectroscopy and fluorescence detectors, labeling OS via reductive amination is widely used to conjugate a chromophore or fluorophore to OS, and acids are necessary to catalyze the reaction [15,16]. One potential issue associated with this strategy is that the acids may partially detach sialic acid monosaccharides from sialylated OS, causing alterations in the structure and function as well as distorting analytical results [17–19]. The labeling step may be circumvented through analyzing native OS with analytical techniques that do not rely on the detection of chromophores or fluorophores, such as high-performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD), high-performance liquid chromatography (HPLC) coupled to refractive index or evaporative light scattering detection, and mass spectrometry. Nonetheless, some steps in the OS analytical workflows may still involve the use of acids. Due to acids' tendency of causing hydrolysis of glycosidic linkages, it is possible that OS analysis steps where acids are necessarily used may result in certain degradation of susceptible OS.

The present work studied the degradation of OS standards during three critical steps in common glycomics analysis workflows, including APTS labeling, solvent evaporation in the presence of TFA, and LC-MS analysis, to examine the susceptibility of specific food OS to acid-induced degradation and-inform about the potential problems of using acids in routine glycomic analysis. We also offer potential solutions to address the undesirable OS degradation and achieve unambiguous OS identification.

2.2. Materials and methods

2.2.1. Materials

Sucrose, raffinose pentahydrate, stachyose hydrate, fructose, glucose, invertase from baker's yeast, and TFA were purchased from MilliporeSigma (St. Louis, MO, USA). 3'-Sialyllactose (3'-SL) and 6'-sialyl-N-acetyllactosamine (6'-SLN) were obtained from V-Labs (now Dextra Laboratories Ltd., Reading, Berkshire, UK). Melibiose was obtained from TCI (Tokyo, Japan). Xylosyl-cellobiose (borohydride reduced) was purchased from Megazyme (Bray, Ireland). N-Glycans were released from an almond flour protein-rich extract by glycopeptidase A (MilliporeSigma).

2.2.2. APTS-labeling of OS and capillary electrophoresis

Glucose, fructose, sucrose, raffinose, stachyose, and melibiose standard solutions (1–2 μL) were transferred to 1.5-mL centrifuge tubes and dried in a centrifugal evaporator (MiVac Quattro, Genevac Ltd., Ipswitch, Suffolk, UK) at room temperature. APTS solution, reductant solution, and strong acid catalyst in a Prozyme GlykoPrep APTS labeling module (Agilent Technologies, Santa Clara, CA, USA) were mixed in the ratio of 1.2:1.2:3.0 (v/v/v). The dried carbohydrate standards were dissolved with 4.5 μL of the mixed labeling reagents and incubated at 65 °C for 1 h in a Thermomixer (Eppendorf North America, Hauppauge, NY, USA). Following incubation, the tubes

were removed from the Thermomixer and were placed in a fume hood with the lid open for 20 min to dissipate any hydrogen cyanide formed during incubation. The samples were mixed with 200 μL of 85% acetonitrile (v/v) and cleaned up with a Gly-Q cartridge module (Agilent Technologies). After cleanup, the samples were diluted and analyzed by the Gly-Q Glycan Analysis System (Agilent Technologies). Brackets of DP 2 and DP 15 (labeled with a non-APTS fluorophore; provided by Agilent Technologies) were added to each injection for aligning electropherograms by glucose units.

2.2.3. Solvent evaporation of OS standards in the presence or absence of TFA

OS standard solutions containing 1 μg of 6'-SLN, 3'-SL, or a mixture of raffinose and stachyose (1 μg each) were prepared in 40% acetonitrile in water (v/v) or 40% acetonitrile in water containing 0.1% TFA (v/v/v) with a total volume of 600 μL. The samples were dried by a centrifugal evaporator at room temperature (no heat applied; ~23–26°C) or 37°C.

2.2.4. Enzymatic treatment with invertase (EC 3.2.1.26) to confirm raffinose and stachyose degradation products

To verify the identities of degradation products generated from raffinose and stachyose during solvent evaporation in the presence of TFA, melibiose and manninotriose were enzymatically produced from raffinose and stachyose standards, respectively, to allow direct comparison with the observed degradation products in the evaporated samples. Invertase is a β-D-fructofuranosidase, which can specifically cleave fructose from raffinose and stachyose, and was chosen for this enzymatic treatment. Raffinose (60 μg of raffinose pentahydrate) and stachyose (80 μg of stachyose hydrate) standards were individually incubated with invertase (10 μg) for 10 min (pH 4.5, 50°C, and 300 rpm) in a Thermomixer. After the incubation, the samples were heated in a boiling water bath for 5 min to inactivate the invertase. The inactivated enzymes were removed

by solid-phase extraction by loading samples onto C18 cartridges (Discovery DSC-18, 100 mg, MilliporeSigma) preconditioned with 2 mL ACN followed by 2 mL water, and then washing the cartridges with 2 mL water to recover carbohydrates. The purified carbohydrates were then analyzed with LC-MS.

2.2.5. LC-MS analysis

LC-MS analysis was performed on an Agilent 6520 Accurate-Mass Q-TOF LC-MS with a Chip Cube interface equipped with an Agilent PGC-Chip II (porous graphitized carbon chip with a 40-nL enrichment column and a 75 μm × 43 mm analytical column). The capillary pump delivered 3% acetonitrile with 0.1% formic acid (v/v/v) at a flow rate of 4 μL min^{-1} and loaded samples onto the enrichment column. The injection volume was 2 μL for each sample. The nano pump delivered mobile phase composed of 3% acetonitrile with 0.1% formic acid (v/v/v) (solvent A) and 89.9% acetonitrile with 0.1% formic acid (v/v/v) (solvent B). The analytes were separated using the nano pump at a flow rate of 0.3 μL min^{-1} with 0% B from 0.0–2.0 min; 0–3% B from 2.0–3.0 min; 3–15% B from 3.0–15.0 min; 15–30% B from 15.0–16.5 min; 30–100% B from 16.5–18.5 min; 100% B 18.5–21.0 min. The mobile phase was switched to 100% A and equilibrated for 9 min before the next injection. The capillary voltage was varied between 1850–1940 V as needed to maintain a stable solvent spray. The drying gas was set at 350 °C with a flow rate of 5 L min^{-1}. When studying the effect of TFA on OS degradation, the dried samples were re-dissolved and diluted with nanopure water to a concentration corresponding to 10 μg mL^{-1} of the original OS and spiked with xylosyl-cellobiose at a concentration of 1 μg mL^{-1} as an internal standard. Peak areas of each analyte were integrated with MassHunter Qualitative Analysis software version B.07.00 (Agilent Technologies) after extracting the protonated molecules ($[M+H]^+$) and corresponding in-source fragment ions and aggregate ions as a merged extracted-ion chromatogram. Relative

quantification was done by normalizing the peak area of analytes against the peak area of xylosyl-cellobiose to compensate for differences in ionization efficiency between runs. These values were then divided by the normalized peak area of the OS samples dried at room temperature in absence of TFA to obtain each OS abundance as a percentage. In a second round of experiments, 0.1% formic acid in the mobile phase was replaced with 5 mM ammonium acetate in an attempt to reduce the extent of in-source fragmentation through the formation of ammonium adducts. Tandem MS analysis was conducted using collision energies determined by the formula [collision energy (V) = 0.013 × *m/z* – 3.5].

2.3. Results and discussion

Two common workflows used in glycomic analysis and the steps at which OS degradation or fragmentation may arise are summarized in Fig. 2.1A and B. Caution is required when these steps are used in glycomic analysis in order to avoid incorrect interpretation of experimental data. Labeling of OS at their reducing end via reductive amination is commonly applied for glycomic analysis to enable or improve the separation and detection of carbohydrates [16,20–23]. Analysis of native (unlabeled) OS is often achieved by mass spectrometry to obtain molecular weights and compositional information, and is an important approach for both targeted and untargeted glycomics [24–26]. The following sections demonstrate and discuss the potential issues of these workflows.

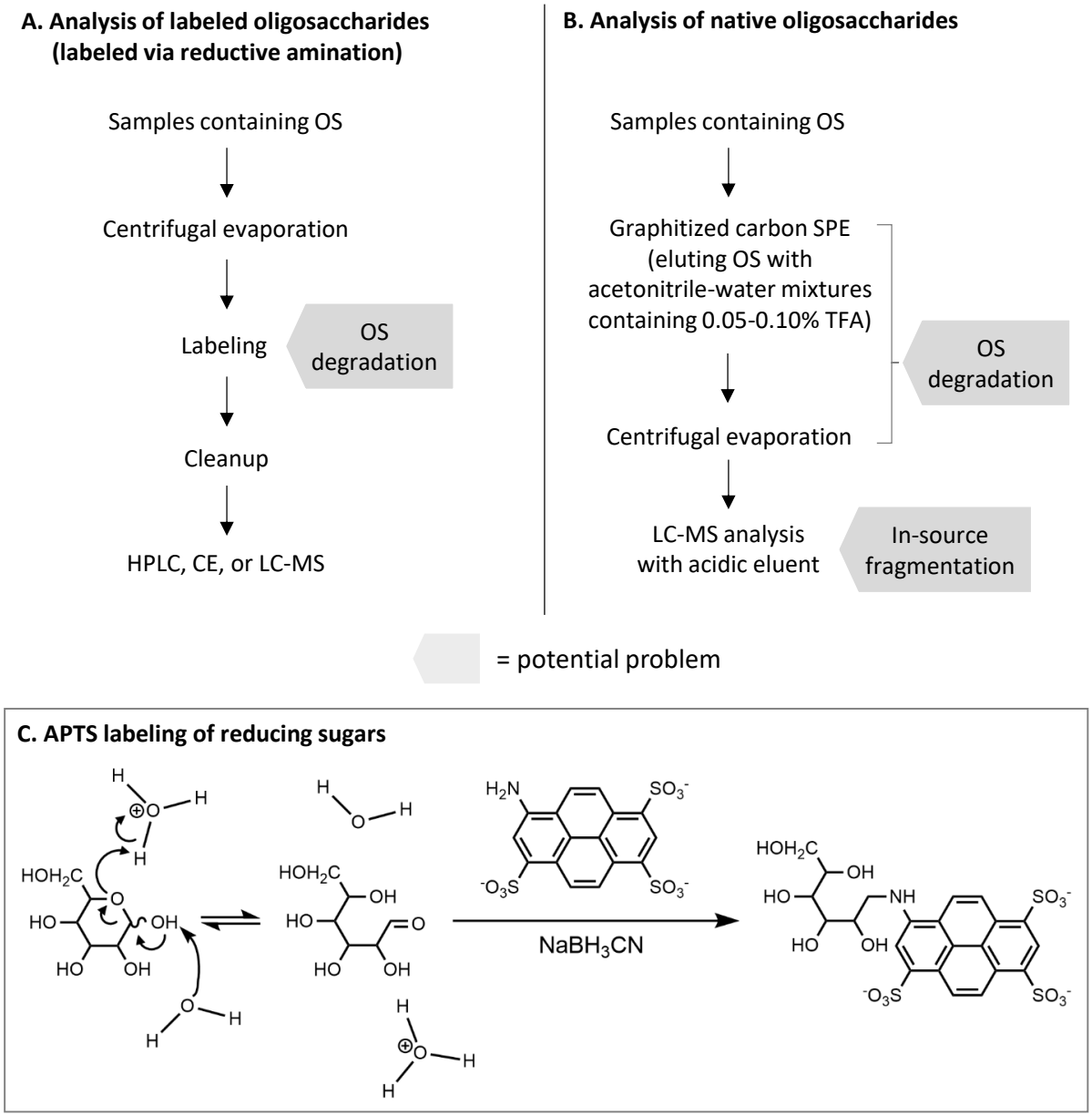


Fig. 2.1. Critical steps in conventional glycomic analysis workflows for labeled (A) and native (B) OS analysis, and APTS labeling reaction to reducing sugars, which are represented by a hexose in this figure (C). The critical steps leading to OS degradation or fragmentation evidenced in the current work are indicated by shaded arrows. CE: capillary electrophoresis.

2.3.1. Degradation of OS during APTS labeling

APTS labeling is one of the OS labeling methods utilizing reductive amination (Fig. 2.1A). To enable the reaction between OS and labeling reagents, a significant amount of acid is typically required to catalyze the hemiacetal ring-opening on the reducing end of OS (Fig. 2.1C). However, it is well known that acids may cause sialic acid monosaccharides to detach from sialylated OS at elevated temperatures (37–60 °C with various incubation times) [17–19]. Therefore, the labeling conditions usually need to be optimized for maximizing the labeling efficiency and minimizing undesirable degradation. In this study, we found that, in addition to the widely known partial degradation of sialylated OS [17–19], some non-reducing sugars might also be susceptible to such acidic labeling conditions. Surprisingly, after undergoing the APTS labeling, sucrose, raffinose, and stachyose displayed clear peaks at 2.4, 3.0, and 3.8 glucose units, respectively, on the CE electropherograms (Fig. 2.2). Because the labeling is based on the reaction between the primary amine of APTS and the aldehyde group on the reducing end of carbohydrates (Fig. 2.1C), non-reducing sugars, such as the three studied here, should neither react with APTS nor generate fluorescence signal after the APTS labeling step. The occurrence of the peaks shown in Fig. 2.2 indicate that sucrose, raffinose, and stachyose were degraded during APTS labeling in a way that exposed free reducing ends. We hypothesized that the α-1,β-2 -glycosidic linkage between glucose and fructose residues was the most labile linkage in sucrose, raffinose, and stachyose. Accordingly, the three peaks observed at 2.4, 3.0, and 3.8 glucose units on the overlaid electropherograms (Fig. 2.2) were assumed to be APTS-labeled glucose, melibiose, and manninotriose, respectively. Fructose, the other degradation product, formed two enantiomers of APTS-labeled fructose that appeared as two small peaks near the peak of the DP 2 bracket (labeled with a manufacturer-proprietary non-APTS fluorophore) as determined by analyzing a fructose standard (data not

shown). Labeling of the degradation product of sucrose and raffinose was further verified by the matching glucose units of APTS-labeled glucose and melibiose standards, respectively (data not shown). Theoretically, any attempt to label non-reducing OS via reductive amination should not produce distinct electropherogram peaks due to the lack of available aldehyde groups, so this result was rather surprising. Importantly, sucrose, raffinose, and stachyose are ubiquitous and abundant in many plant foods, including legumes, peanuts, and tree nuts, and are often analyzed during routine analysis and quality control of food products. Therefore, one must consider that when analyzing OS (e.g., free OS, released N-glycans, and OS derived from polysaccharides) labeled via reductive amination in complex plant-derived samples, the peaks resulting from cleavage of non-reducing carbohydrates may lead to misidentification and inaccurate characterization/quantification. Hence, analysts should be aware of this issue when analyzing samples containing such susceptible OS. Whenever possible, the OS of interest should be separated from non-reducing carbohydrates before labeling, such as through gel filtration. In particular, for the analysis of N-glycans, protein precipitation or membrane filtration could be performed before releasing N-glycans in order to exclude non-reducing soluble carbohydrates.

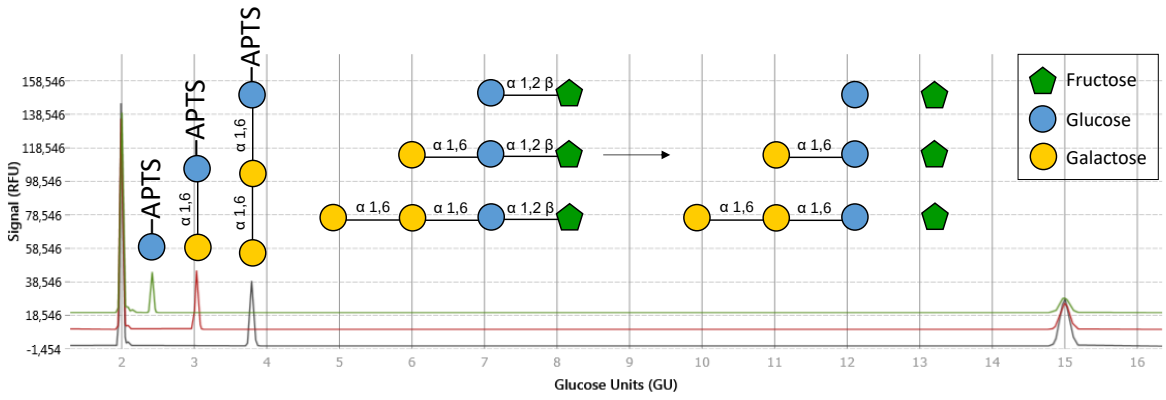
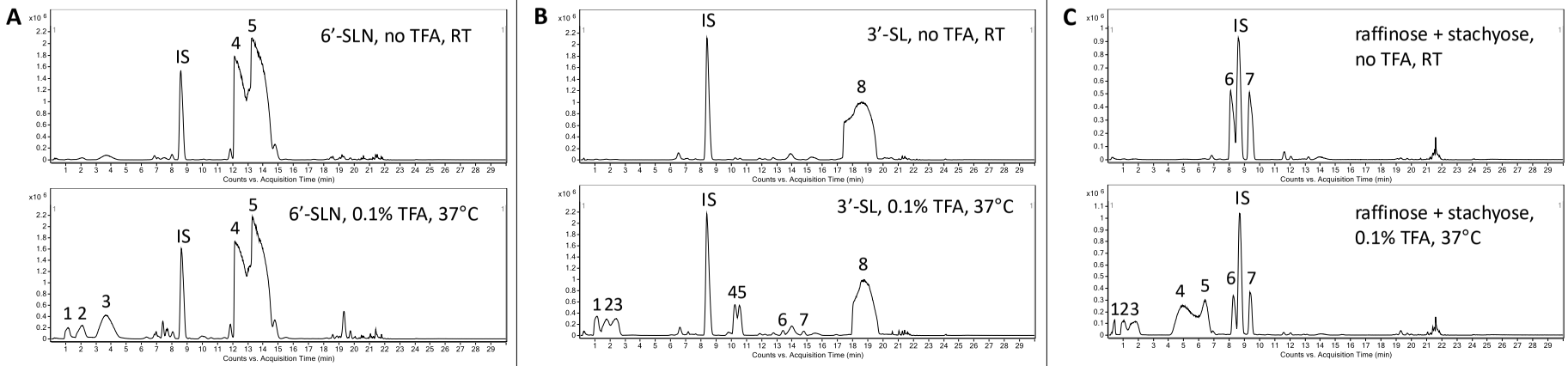


Fig. 2.2. Overlaid Gly-Q capillary electropherograms of degradation products of sucrose (top), raffinose (middle), and stachyose (bottom) following APTS labeling treatment at 65 °C for 1 h. Peaks at 2 and 15 glucose units (GU) are DP 2 and DP 15 brackets, respectively.

2.3.2. Effect of acid on OS degradation during solvent evaporation at mild temperature

Graphitized carbon is the method of choice for desalting OS in food samples or biological matrices prior to analysis by LC-MS (Fig. 2.1B) [27,28]. Neutral and acidic OS can be sequentially eluted from a graphitized carbon column with acetonitrile-water mixtures and acetonitrile-water mixtures containing dilute TFA; they can also be directly eluted as a single fraction with acetonitrile-water mixtures containing dilute TFA [27,28]. Although lyophilization can fulfill the subsequent step of solvent removal, centrifugal evaporation is more frequently used as it is more affordable and time-efficient. In this study, we tested four standard OS, 6'-SLN and 3'-SL (milk OS) and raffinose and stachyose (legume OS), during solvent evaporation and discovered significant degradation. After drying the OS solutions at 37 °C to remove the solvent containing 0.1% TFA, LC-MS analysis revealed additional peaks in the chromatograms corresponding to degradation products (Fig. 2.3). The chromatogram of 6'-SLN displayed several additional peaks generated by degradation that were not present when 6'-SLN was dried in absence of acid at room

temperature (Fig. 2.3A). The new peaks were confirmed by mass as 6'-SLN degradation products: N-acetylneuraminic acid (Neu5Ac, peak 1) and two anomers of N-acetyllactosamine (peaks 2 and 3). Similarly, Neu5Ac (peak 1) and two anomers of lactose (peaks 2 and 3) were identified in the chromatogram generated from 3'-SL dried in solvent containing 0.1% TFA (Fig. 2.3B). In addition to the products of desialylation, 3'-SL generated four additional peaks (peaks 4, 5, 6, and 7) when drying the OS in 40% acetonitrile with 0.1% TFA. These four peaks were all chiefly composed of a mixture of two ions with *m/z* 616.21 and 598.20, which could be the products of acid-catalyzed lactonization involving the carboxyl group on the Neu5Ac residue and one of the hydroxyl groups on galactose residue [29,30]. Degradation of 6'-SLN and 3'-SL was also evident in the abundance of each OS measured after solvent evaporation. To obtain the relative abundances of the OS dried under each set of conditions, the peak areas of each OS were first normalized against the peak area of the internal standard to compensate for differences in ionization efficiency. After normalizing against the internal standard, the samples were normalized against the data collected for the samples dried at room temperature in absence of TFA, which were considered the control group. In-source fragment ions and aggregate ions were merged with the protonated molecules ($[M+H]^+$) of the OS for peak area integration. More details about in-source fragmentation and aggregate ion formation are discussed in the next section. The relative abundances of 6'-SLN and 3'-SL standards that underwent evaporation in 40% acetonitrile with 0.1% TFA at 37 °C were 8.5% and 26.5% lower, respectively, than when dried at room temperature in the absence of TFA (Fig. 2.4A and B).



Peak no.	RT (min)	MS1 major peak m/z	Compound
1	1.1	**310.11**, 292.10, 619.22	Sialic acid
2	2.1	**384.15**, 366.14, 767.29	N-Acetyllactosamine
3	3.6	**384.15**, 366.14, 767.29	N-Acetyllactosamine
4	12.1	**675.25**	6'-SLN
5	13.4	**675.25**	6'-SLN

Peak no.	RT (min)	MS1 major peak m/z	(Suggested) compound
1	1.1	**310.11**, 619.22, 292.10	Sialic acid
2	1.8	685.24, 325.11, **343.12**, 163.06	Lactose
3	2.4	685.24, **343.12**, 325.11, 163.06	Lactose
4	10.2	**616.21**, 598.20	3'-SL lactone
5	10.5	**616.21**, 598.20	3'-SL lactone
6	13.4	598.20, **616.21**	3'-SL lactone
7	14.8	**616.21**, 598.20	3'-SL lactone
8	18.7	454.16, **634.22**, 292.10	3'-SL

Peak no.	RT (min)	MS1 major peak m/z	Compound
1	0.4	163.06, **325.11**	Fructose
2	1.1	163.06, 325.11, 685.24, 667.23, **343.12**	Melibiose
3	1.8	163.06, 325.11, 685.24, 667.23, **343.12**	Melibiose
4	5.0	487.17, 325.11, **505.18**, 163.06, 343.12	Manninotriose
5	6.4	487.17, 325.11, **505.18**, 163.06, 343.12	Manninotriose
6	8.1	325.11, 163.06, 343.12, 685.24, 847.29	Raffinose
7	9.4	505.18, 487.17, 325.11, 163.06, 343.12	Stachyose

Fig. 2.3. LC-Q-TOF base peak chromatograms of 6'-SLN (A), 3'-SL (B), and a mixture of raffinose and stachyose (C) prepared in 40% acetonitrile and dried at room temperature (no TFA, top chromatogram), and prepared in 40% acetonitrile containing 0.1% TFA and dried at 37 °C (bottom chromatogram). Xylosyl-cellobiose (internal standard, IS) was added to each sample at a concentration of 1 μg mL^{-1} before injection. The major peaks in the ESI mass spectra of each chromatographic peak are summarized in the tables (ranked by signal intensity from high to low) below each pair of chromatograms. The m/z values of the protonated molecules ($[M+H]^+$) are marked in bold.

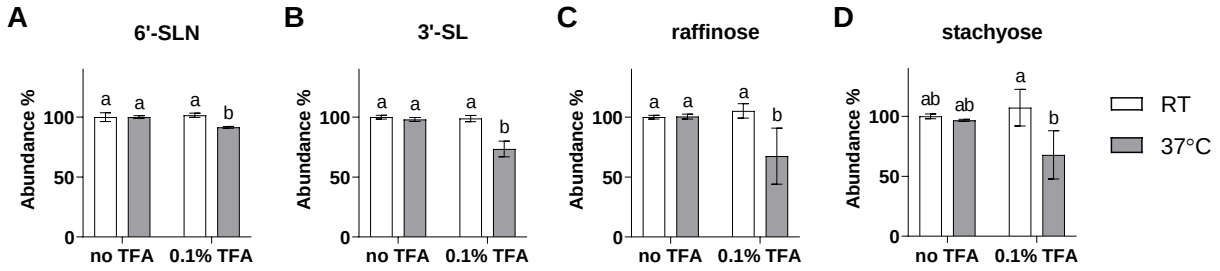


Fig. 2.4. Relative abundances of 3'-SL (A), 6'-SLN (B), raffinose (C), and stachyose (D) after evaporation of different solvents (40% acetonitrile with or without 0.1% TFA) at various temperatures (room temperature (RT) or 37 °C). Relative abundances were measured by LC-Q-TOF and are normalized to the samples dried in absence of TFA at room temperature. Data are presented as mean ± standard deviation. Different letters on the bars indicate significant differences for each analyte among samples subjected to different treatments ($p < 0.05$ by Tukey's method).

For raffinose and stachyose, the additional chromatographic peaks generated during the evaporation of 40% acetonitrile with 0.1% TFA (Fig. 2.3C) were tentatively identified as fructose (peak 1), two anomers of melibiose (peaks 2 and 3), and two anomers of manninotriose (peaks 4 and 5). Injecting raffinose and stachyose samples separately confirmed that peaks 2 and 3 were generated from raffinose, and peaks 4 and 5 were from stachyose. The monosaccharides composing raffinose and stachyose (fructose, glucose, and galactose) are all hexoses, and thus the resulting degradation products could not be unambiguously identified by mass spectrometry because the monosaccharides have identical masses. To verify our hypothesis that raffinose and stachyose generated melibiose and manninotriose during solvent evaporation, we conducted an enzymatic treatment using the enzyme invertase on raffinose and stachyose. The retention times of the enzymatically generated melibiose and manninotriose matched precisely the degradation

products generated from raffinose and stachyose during solvent evaporation at 37 °C in the presence of TFA. The relative abundances of raffinose and stachyose were lower in the samples dried at 37 °C in presence of TFA compared with the samples without acid and/or dried at room temperature (Fig. 2.4C and D). As there was no significant loss for any of the four OS when dried at room temperature in 40% acetonitrile with 0.1% TFA (Fig. 2.4A–D), we proposed that when removing solvents containing TFA from OS samples by centrifugal evaporation, no heating should be applied. While it is well known that concentrated TFA is often used for breaking down OS or polysaccharides at high temperatures for subsequent monosaccharide composition analysis [12,13], this work demonstrated that even minute TFA concentrations at mild temperatures can cause partial degradation of susceptible OS. Hence, this effect needs to be taken into account when using TFA, even dilute, in glycomic analysis.

2.3.3. Evaluating in-source fragmentation of OS in LC-MS analysis

When analyzing OS with LC-MS in the positive ion mode, in most cases, $[M+H]^+$ or $[M+H–H_2O]^+$ represent the major peaks in the ESI mass spectra. However, this study revealed significant in-source fragmentation for many OS, including a representative OS found in bovine milk (3'-SL) and two important OS (raffinose and stachyose) abundant in legumes (Fig. 2.5). The tallest peak in the ESI mass spectra of 3'-SL was *m/z* 454.16 ($[M−Hex−H_2O+H]^+$), followed by the peaks *m/z* 634.22 ($[M+H]^+$), *m/z* 292.10 (($[M−2Hex−H_2O+H]^+$ or $[Neu5Ac−H_2O+H]^+$), and *m/z* 1267.43 ($[2M+H]^+$) (Fig. 2.5A). By comparison, the MS1 spectra of raffinose and stachyose were even more complex. For raffinose, the most abundant peaks in the ESI mass spectra, ranked by signal intensity from high to low, were *m/z* 325.11, *m/z* 163.06, *m/z* 343.12, and *m/z* 685.24 (Fig. 2.5B). The theoretical *m/z* of protonated raffinose, 505.18 ($[M+H]^+$), was much less abundant than the other ions mentioned above. Similarly, the tallest peak in the ESI mass spectra of stachyose did

not correspond to the $[M+H]^+$, but rather to its fragments. The most abundant MS1 ions of stachyose were *m/z* 505.18, *m/z* 487.17, and *m/z* 325.11, whereas the theoretical *m/z* of stachyose, 667.23 ($[M+H]^+$) and 649.22 ($[M-H_2O+H]^+$), were found in extremely low abundance (Fig. 2.5C). Therefore, if one were unaware of this potential for in-source fragmentation and were to make OS assignments only based on the observed ESI mass spectra, it is highly possible that a fragment of stachyose would be mis-annotated as an OS consisting of three hexose residues. Similar issues were also recently reported for LC-MS-based cellular metabolomics [31] and MS-based lipidomics [32].

In-source fragmentation of OS in LC-MS analysis is usually not severe and only slightly decreases the signal intensities of protonated molecules. However, we observed that the α-1,β-2 -glycosidic linkage between the glucose and fructose in raffinose and stachyose was extraordinarily labile compared with other common glycosidic linkages, similar to our prior observation with the APTS labeling in acid and solvent evaporation in presence of TFA at 37°C. For the LC-MS analysis of raffinose, we suggest that the α-1,β-2 -glycosidic linkage was cleaved either at the electrospray ionization (ESI) stage or after ESI but before the ions entered the mass analyzer, thus abundant fructose (*m/z* 163.06) and melibiose (*m/z* 325.11 and *m/z* 343.12) were generated (Fig. 2.5B). Moreover, just after the fragmentation, two melibiose units formed an aggregate ion $[2M+H]^+$ to generate the peak at *m/z* 685.24. In the MS1 spectra of stachyose (Fig. 2.5C), the predominant peaks at *m/z* 505.18 and *m/z* 487.17 were identified as manninotriose, one of the two products of α-1,β-2 linkage cleavage from stachyose. The significant peak at *m/z* 325.11 indicated that in-source fragmentation also occurred on the α-1,6-glycosidic linkage in the middle of stachyose to generate OS containing two hexose residues, whereas the peaks at *m/z* 1171.40 and *m/z* 1009.35 indicated the occurrence of aggregate ion formation where stachyose-manninotriose

aggregate ions and manninotriose-manninotriose aggregate ions, respectively, were formed. In-source fragmentation and aggregate ion formation have occasionally been seen in the analysis of OS using ESI-MS [33–35], but to our knowledge, aggregate ion formation involving in-source fragmentation products was not reported previously. In previous studies employing LC-ESI-MS for the analysis of raffinose and stachyose, $[M+Na]^+$ (*m/z* 527 and *m/z* 689, respectively), $[M-H]^-$ (*m/z* 503 and *m/z* 665, respectively), or $[M+HCOO]^-$ (*m/z* 549 and *m/z* 771, respectively) were the most significant ions in the ESI mass spectra [36–38]. Because the OS ions in the current study were all protonated, we suspect that protonation, which was reported to facilitate glycosidic linkage fragmentation previously [39,40], contributed substantially to the cleavage of α-1,β-2 -glycosidic linkage in raffinose and stachyose.

Aggregate ion formation was also observed with melibiose and manninotriose, which formed aggregate ions of *m/z* 685.24 (Supporting information Fig. 2.S1A and C) and *m/z* 1009.3446 (Supporting information Fig. 2.S1B and D), respectively, in the ESI mass spectra. This further corroborated our hypothesis that the ions *m/z* 685.24 and *m/z* 1009.35 in the ESI mass spectra of raffinose and stachyose, respectively, were in reality aggregate ions formed by the in-source fragments.

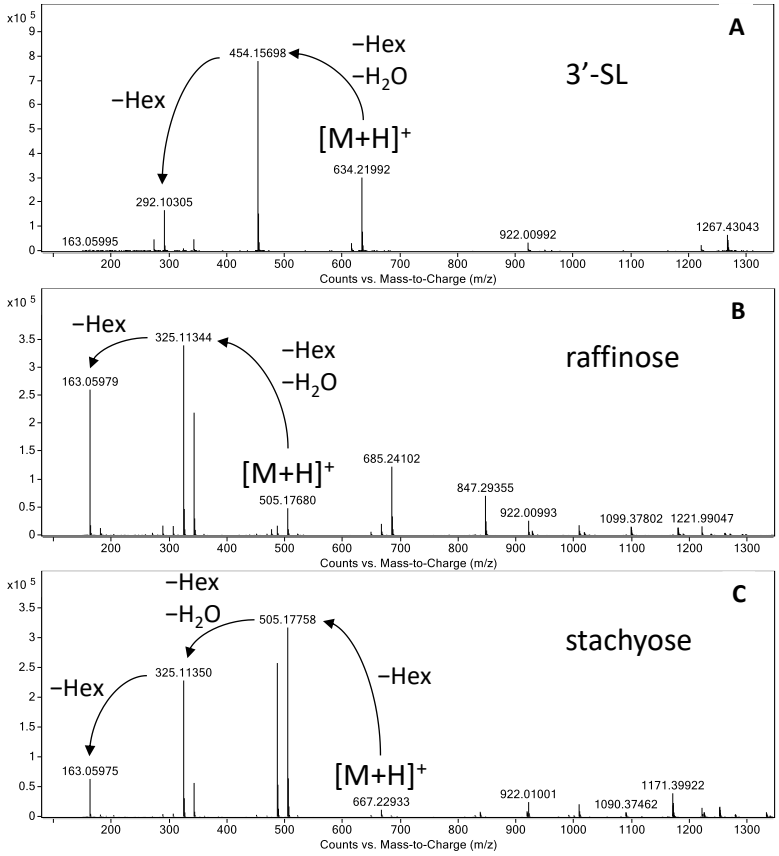


Fig. 2.5. LC-Q-TOF ESI mass spectra of standards of 3'-SL (A), raffinose (B), and stachyose (C) with extensive in-source fragmentation. Peaks 922.01 and 1221.99 are calibration ions.

To further confirm that the ions observed in the MS1 spectra originated from the same OS (3'-SL, raffinose, or stachyose), we extracted and overlaid the chromatograms of the major ions (Fig. 6). The retention times of all the major ions in the MS1 spectra of 3'-SL (*m/z* 454.16, *m/z* 634.22, *m/z* 292.10, *m/z* 1267.43, and *m/z* 343.12), raffinose (*m/z* 325.11, *m/z* 163.06, *m/z* 343.12, *m/z* 685.24, *m/z* 847.29, *m/z* 505.18, *m/z* 667.23, *m/z* 1009.35, and *m/z* 487.17), and stachyose (*m/z* 505.18, *m/z* 487.17, *m/z* 325.11, *m/z* 163.06, *m/z* 343.12, *m/z* 1171.40, *m/z* 1009.35, and *m/z* 667.23), respectively, aligned correctly, providing evidence that those ions were indeed generated by in-source fragmentation and aggregate ion formation. The chromatographic peaks among

different extracted-ion chromatograms (EIC) were also well aligned when the released N-glycans (Supporting information Fig. 2.S2) and galactooligosaccharides were analyzed. Therefore, while analyzing real samples, overlaying the EIC of all identified OS could assist in identifying in-source fragment ions and prevent mis-annotation of the fragments as genuine OS [41].

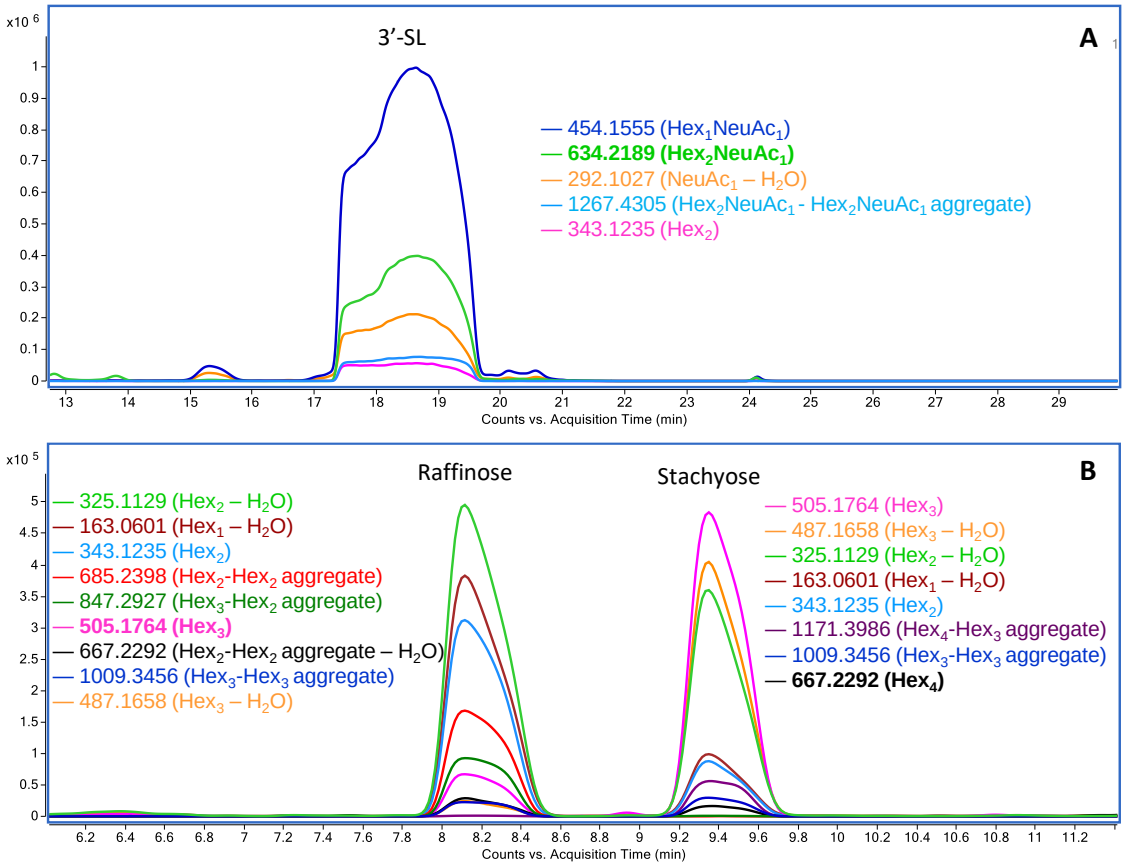


Fig. 2.6. Overlaid LC-Q-TOF chromatograms (EIC) of protonated molecules ($[M+H]^+$), in-source fragment ions, and aggregate ions of 3'-SL (A), and raffinose and stachyose (B). The m/z values of the protonated molecules are marked in bold. Hex: hexose. NeuAc: N-acetylneuraminic acid.

2.3.4. Solutions for avoiding mis-annotation of in-source fragments as genuine OS in LC-MS analysis

The degree of in-source fragmentation can be affected by the capillary voltage, drying gas temperature, and numerous other parameters of the mass spectrometer. Despite adjusting those parameters, raffinose and stachyose still had extensive in-source fragmentation (data not shown). Although negative ion mode and permethylation may stabilize the labile sialylated OS (and are sometimes used) [42,43], analyzing OS in native forms in positive ion mode is still extensively done for both targeted and untargeted glycomics [24,44–46]. Even if the majority of OS may not be severely affected, being aware of the potential in-source fragmentation of particular OS and the consequential effects is crucial to correctly interpret the LC-MS data (Fig. 2.1B).

As mentioned above, protonated molecules were barely observed in the ESI mass spectra of some analytes, such as stachyose. Thus, correct assignment of OS identities (i.e., DP and monosaccharide composition) without performing a separate comparison with the corresponding standards would not be possible. In particular, when performing untargeted analysis for discovering novel OS in foods, it is usually not possible to tell whether the identified OS contain fragile linkages or whether a particular mass spectral peak represents a true OS or an in-source fragment. Therefore, it is crucial to find a strategy to deal with the potential incorrect identifications that may result from extensive in-source fragmentation.

As we suspected protons would initiate in-source fragmentation of susceptible OS, we tried to replace formic acid with other mobile phase additives, such as ammonium acetate, to reduce the proton concentration and facilitate the formation of adducts other than protonated molecules. When a mobile phase containing 5 mM ammonium acetate was used, in-source fragmentation of raffinose and stachyose was greatly diminished (Fig. 2.7A, B). In their ESI mass spectra,

ammonium adduct ions were the most intense peaks, and sodium and potassium adduct ions were also observed at lower intensities. For 3'-SL, in-source fragmentation was also reduced when changing the mobile phase additive from 0.1% formic acid to 5 mM ammonium acetate (Fig. 2.7C). However, the protonated form of 3'-SL was still more abundant than other adduct ions ($[M+NH_4]^+$, $[M+Na]^+$, and $[M+K]^+$), and there was still a considerable abundance of in-source fragment ions (*m/z* 454.16). The relative abundance of protonated molecules and the other adduct ions of different OS may be related to the specific OS molecule properties, such as elemental composition [47].

The enhanced stability of the ammonium and metal ion adducts was also revealed in their tandem-MS fragmentation behavior. For example, when using collision energy settings that were optimized for protonated carbohydrates, the abundance of the protonated raffinose precursor in its MS2 spectra was lower than the abundance of ammonium adduct ions in their MS2 spectra, and sodium and potassium adduct ions remained completely unfragmented (Supporting information Fig. 2.S3). This confirmed that protonated ions were less stable and fragmented more easily compared with the other adducts. An alkaline metal ion can coordinate with several oxygen atoms in an OS molecule and thus stabilize the adduct ions, while a proton can coordinate with at most two oxygen atoms [40]. Ammonium ions appear to behave similarly to alkaline metal ions, as the ammonium adducts were also more stable than the protonated molecules.

Here we identified a simple strategy that can reduce the probability of incorrect oligosaccharide identification resulting from extensive in-source fragmentation. In the ESI mass spectra of raffinose, stachyose, and 3'-SL analyzed with a mobile phase containing 5 mM ammonium acetate, the identity of the true intact analyte is evidenced by the existence of $[M+NH_4]^+$, $[M+Na]^+$, and $[M+K]^+$, while the in-source fragments did not produce these adducts.

Therefore, when conducting untargeted analysis, the presence of these adduct ions in significant abundances would suggest the authenticity of the OS. In summary, modifying the mobile phase to prevent the formation of labile protonated molecules could reduce in-source fragmentation and, more importantly, ensure correct identification of susceptible OS.

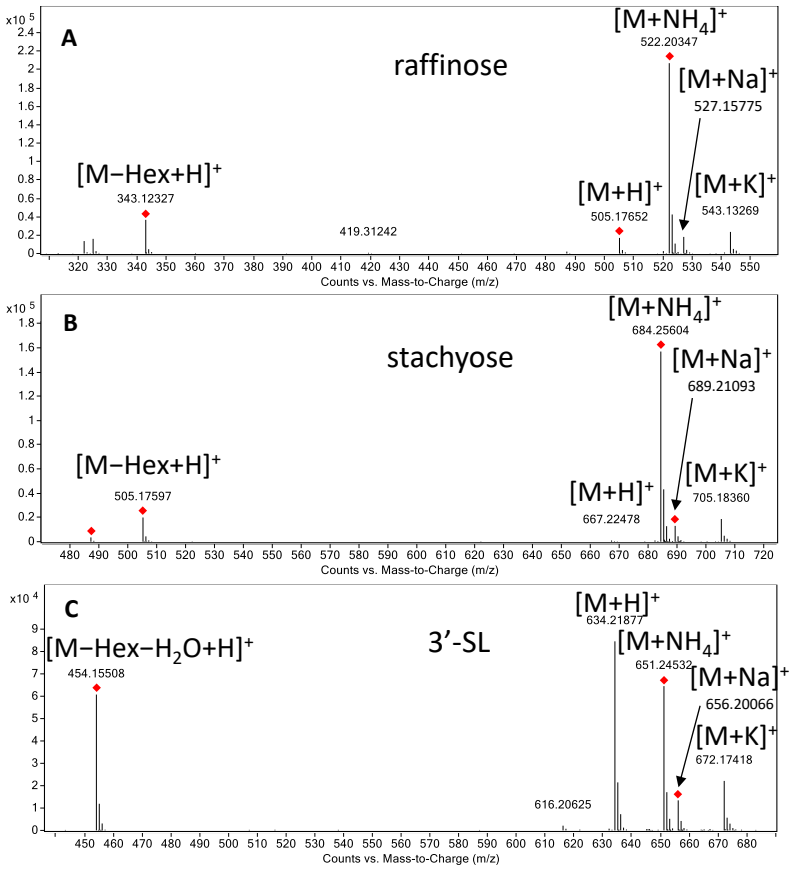


Fig. 2.7. LC-Q-TOF ESI mass spectra of raffinose (A), stachyose (B), and 3'-SL (C) run with a mobile phase containing 5 mM ammonium acetate in positive ion mode.

2.4. Conclusions

This study demonstrates the susceptibility of specific OS linkages to degradation induced by heat and acid. In addition to the well-known susceptible linkages of sialylation, the α-1,β-2

linkages in sucrose, raffinose, and stachyose are also labile in an acidic environment. This degradation occurred not only at high temperatures in presence of concentrated acids but also at mild temperatures in the presence of dilute acids, as commonly used in OS purification workflows. Therefore, careful consideration must be given to ensure the accuracy of glycan characterization when acids are used at any step in a conventional glycomics workflow. This is especially important for samples containing OS consisting of sialyl linkages or α-1,β-2 glycosidic linkages. To eliminate the potential interference caused by the degradation products of non-reducing OS during labeling and analysis, it is advisable to fractionate such susceptible non-reducing sugars before labeling. The degradation of OS during centrifugal evaporation of native OS can be prevented by drying samples at room temperature. In-source fragmentation in LC-MS analysis under positive ion mode can be greatly diminished by changing the mobile phase additive from formic acid to ammonium acetate to reduce the formation of labile protonated molecules. Incorrect identification for the susceptible OS can be successfully avoided by distinguishing authentic OS from in-source fragment ions with the presence of ammonium, sodium, and potassium adduct ions.

Acknowledgements

The authors thank Prozyme Inc. (now Agilent) for providing the APTS-labeling reagents and Gly-Q consumables. This study was funded by the U.S. Department of Agriculture's (USDA) Agricultural Marketing Service through grant AM170100XXXXG011 and USDA NIFA Hatch grant CA-D-FST-2187-H. The content of this publication is solely the responsibility of the authors and does not necessarily represent the official views of the USDA. Yu-Ping Huang is a recipient of Yen Chuang Taiwan Fellowship.

Supporting information

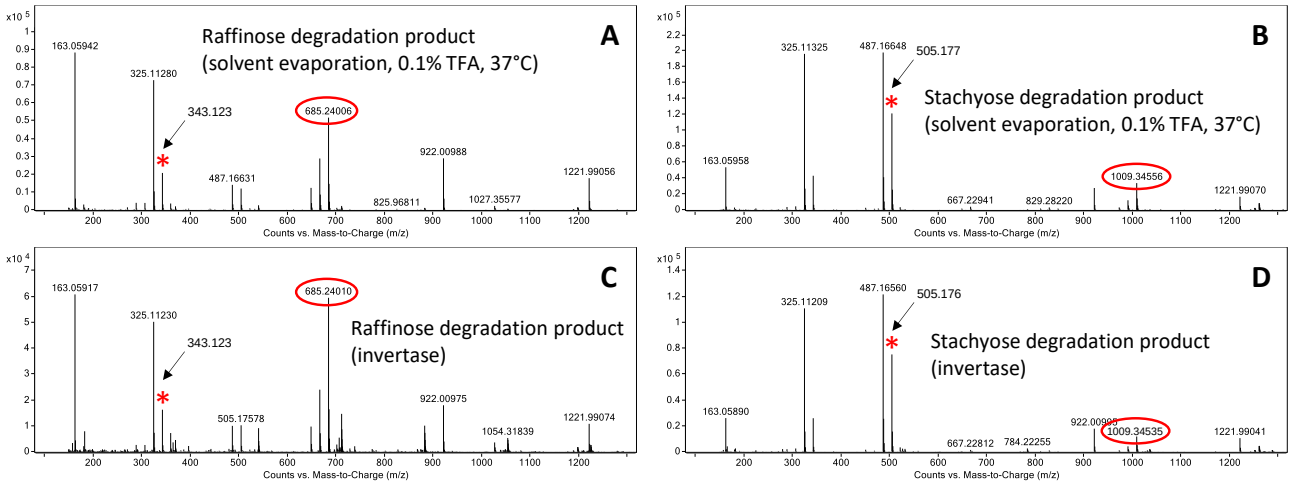


Fig. 2.S1. LC-Q-TOF ESI mass spectra of degradation products at 1.8 min (A) and 6.4 min (B) generated from raffinose and stachyose, respectively, during solvent evaporation at 37°C in the presence of TFA, and the spectra of melibiose (C) and manninotriose (D) generated through invertase enzymatic treatment of raffinose and stachyose, respectively. Peaks denoted by asterisks are the protonated molecules ($[M+H]^+$). The m/z values of aggregate ions are marked in red circles. Peaks 922.01 and 1221.99 are calibration ions.

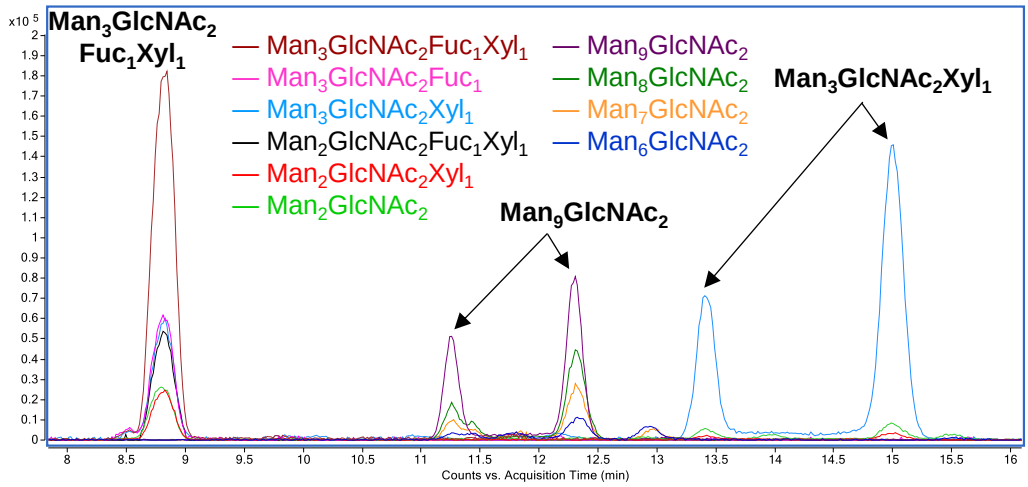


Fig. 2.S2. Overlaid chromatograms (EIC) of protonated molecules and the corresponding in-source fragment ions of selected N-glycans ($Man_3GlcNAc_2Fuc_1Xyl_1$, $Man_9GlcNAc_2$, and $Man_3GlcNAc_2Xyl_1$) released from almond proteins. Man: mannose. GlcNAc: N-Acetylglucosamine. Fuc: fucose. Xyl: xylose.

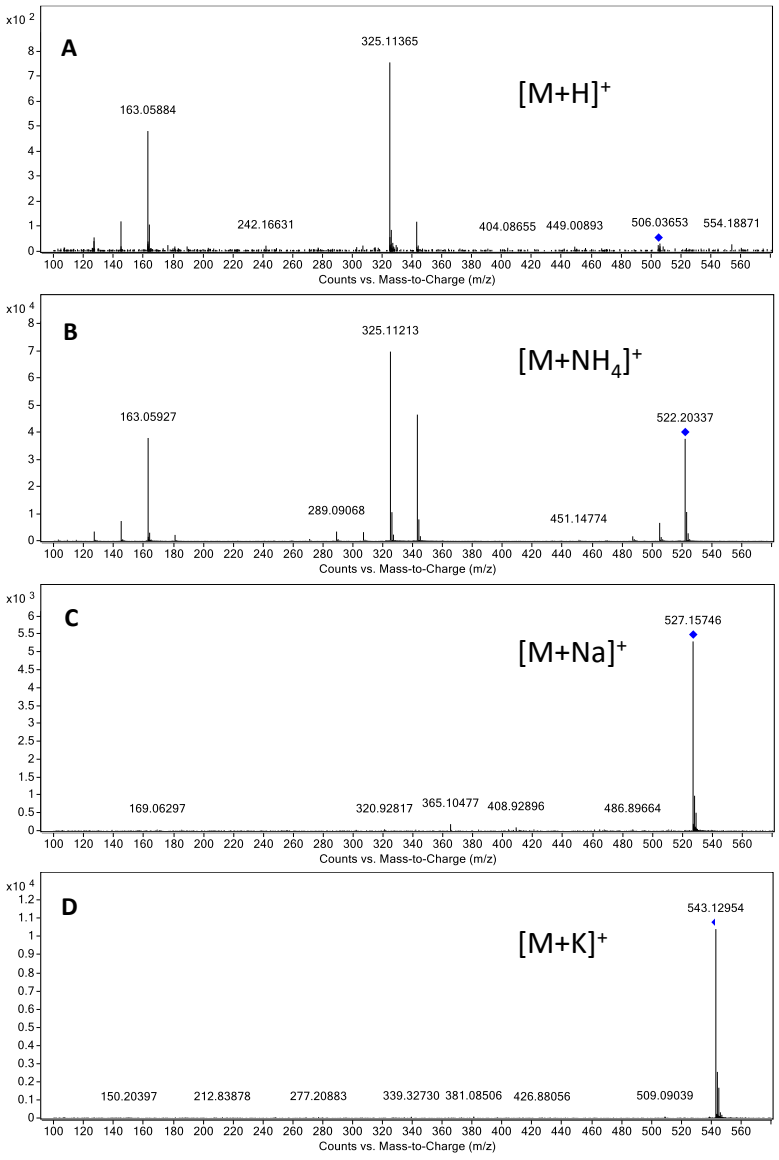


Fig. 2.S3. LC-Q-TOF MS2 spectra of $[M+H]^+$ (collision energy 3.07 V; A), $[M+NH_4]^+$ (collision energy 3.29 V; B), $[M+Na]^+$ (collision energy 3.35 V; C), and $[M+K]^+$ (collision energy 3,56 V; D) of raffinose.

References

[1] Patel S, Goyal A. Functional oligosaccharides: production, properties and applications. World J Microbiol Biotechnol 2011;27:1119–28. https://doi.org/10.1007/s11274-010-0558-5.

[2] Parc AL, Karav S, Leite Nobrega de Moura Bell JM, Frese SA, Liu Y, Mills DA, et al. A novel endo-β-N-acetylglucosaminidase releases specific N-glycans depending on different reaction conditions. Biotechnol Prog 2015;31:1323–30. https://doi.org/10.1002/btpr.2133.

[3] Mussatto SI, Mancilha IM. Non-digestible oligosaccharides: A review. Carbohydr Polym 2007;68:587–97. https://doi.org/10.1016/j.carbpol.2006.12.011.

[4] Lai H-M, Huang Y-P. Effects of Ball Milling on the Properties of Colored Rice Bran. Gums and Stabilisers for the Food Industry 17, 2014, p. 155–63. https://doi.org/10.1039/9781782621300-00155.

[5] Parajó JC, Garrote G, Cruz JM, Dominguez H. Production of xylooligosaccharides by autohydrolysis of lignocellulosic materials. Trends in Food Science & Technology 2004;15:115–20. https://doi.org/10.1016/j.tifs.2003.09.009.

[6] Xing H, Yaylayan VA. Mechanochemical depolymerization of inulin. Carbohydr Res 2018;460:14–8. https://doi.org/10.1016/j.carres.2018.02.009.

[7] Yuan X, Wang J, Yao H. Production of feruloyl oligosaccharides from wheat bran insoluble dietary fibre by xylanases from Bacillus subtilis. Food Chem 2006;95:484–92. https://doi.org/10.1016/j.foodchem.2005.01.043.

[8] Bunyatratchata A, Huang Y-P, Ozturk G, Cohen JL, Bhattacharya M, MLN de Moura Bell J, et al. Effects of Industrial Thermal Treatments on the Release of Bovine Colostrum Glycoprotein N-Glycans by Endo-β-N-acetylglucosaminidase. J Agric Food Chem 2020;68:15208–15. https://doi.org/10.1021/acs.jafc.0c05986.

[9] Swennen K, Courtin CM, Delcour JA. Non-digestible Oligosaccharides with Prebiotic Properties. Crit Rev Food Sci Nutr 2006;46:459–71. https://doi.org/10.1080/10408390500215746.

[10] Moura FA de, Macagnan FT, Silva LP da. Oligosaccharide production by hydrolysis of polysaccharides: a review. Int J Food Sci 2015;50:275–81. https://doi.org/10.1111/ijfs.12681.

[11] Wang S, Copeland L. Effect of Acid Hydrolysis on Starch Structure and Functionality: A Review. Crit Rev Food Sci Nutr 2015;55:1081–97. https://doi.org/10.1080/10408398.2012.684551.

[12] Shi H, Wan Y, Li O, Zhang X, Xie M, Nie S, et al. Two-step hydrolysis method for monosaccharide composition analysis of natural polysaccharides rich in uronic acids. Food Hydrocoll 2020;101:105524. https://doi.org/10.1016/j.foodhyd.2019.105524.

[13] Tabarsa M, You S, Dabaghian EH, Surayot U. Water-soluble polysaccharides from Ulva intestinalis: Molecular properties, structural elucidation and immunomodulatory activities. J Food Drug Anal 2018;26:599–608. https://doi.org/10.1016/j.jfda.2017.07.016.

[14] Miller RL, Guimond SE, Prescott M, Turnbull JE, Karlsson N. Versatile Separation and Analysis of Heparan Sulfate Oligosaccharides Using Graphitized Carbon Liquid Chromatography and Electrospray Mass Spectrometry. Anal Chem 2017;89:8942–50. https://doi.org/10.1021/acs.analchem.7b01417.

[15] Lamari FN, Kuhn R, Karamanos NK. Derivatization of carbohydrates for chromatographic, electrophoretic and mass spectrometric structure analysis. J Chromatogr B 2003;793:15–36. https://doi.org/10.1016/S1570-0232(03)00362-3.

[16] Ruhaak LR, Xu G, Li Q, Goonatilleke E, Lebrilla CB. Mass Spectrometry Approaches to Glycomic and Glycoproteomic Analyses. Chem Rev 2018;118:7886–930. https://doi.org/10.1021/acs.chemrev.7b00732.

[17] Aich U, Hurum DC, Basumallick L, Rao S, Pohl C, Rohrer JS, et al. Evaluation of desialylation during 2-amino benzamide labeling of asparagine-linked oligosaccharides. Anal Biochem 2014;458:27–36. https://doi.org/10.1016/j.ab.2014.03.008.

[18] Bigge JC, Patel TP, Bruce JA, Goulding PN, Charles SM, Parekh RB. Nonselective and Efficient Fluorescent Labeling of Glycans Using 2-Amino Benzamide and Anthranilic Acid. Anal Biochem 1995;230:229–38. https://doi.org/10.1006/abio.1995.1468.

[19] Evangelista RA, Chen F-TA, Guttman A. Reductive amination of N-linked oligosaccharides using organic acid catalysts. J Chromatogr A 1996;745:273–80. https://doi.org/10.1016/0021-9673(96)00266-X.

[20] Albrecht S, Schols HA, van Zoeren D, van Lingen RA, Groot Jebbink LJM, van den Heuvel EGHM, et al. Oligosaccharides in feces of breast- and formula-fed babies. Carbohydr Res 2011;346:2173–81. https://doi.org/10.1016/j.carres.2011.06.034.

[21] Galeotti F, Coppa GV, Zampini L, Maccari F, Galeazzi T, Padella L, et al. Capillary electrophoresis separation of human milk neutral and acidic oligosaccharides derivatized with 2-aminoacridone. Electrophoresis 2014;35:811–8. https://doi.org/10.1002/elps.201300490.

[22] Lorenz D, Erasmy N, Akil Y, Saake B. A new method for the quantification of monosaccharides, uronic acids and oligosaccharides in partially hydrolyzed xylans by HPAEC-UV/VIS. Carbohydr Polym 2016;140:181–7. https://doi.org/10.1016/j.carbpol.2015.12.027.

[23] Ruhaak LR, Steenvoorden E, Koeleman CAM, Deelder AM, Wuhrer M. 2-Picoline-borane: A non-toxic reducing agent for oligosaccharide labeling by reductive amination. Proteomics 2010;10:2330–6. https://doi.org/10.1002/pmic.200900804.

[24] Aldredge DL, Geronimo MR, Hua S, Nwosu CC, Lebrilla CB, Barile D. Annotation and structural elucidation of bovine milk oligosaccharides and determination of novel fucosylated structures. Glycobiology 2013;23:664–76. https://doi.org/10.1093/glycob/cwt007.

[25] Karav S, Parc AL, Bell JMLN de M, Frese SA, Kirmiz N, Block DE, et al. Oligosaccharides Released from Milk Glycoproteins Are Selective Growth Substrates for Infant-Associated Bifidobacteria. Appl Environ Microbiol 2016;82:3622–30. https://doi.org/10.1128/AEM.00547-16.

[26] Tian T, Freeman S, Corey M, German JB, Barile D. Chemical Characterization of Potentially Prebiotic Oligosaccharides in Brewed Coffee and Spent Coffee Grounds. J Agric Food Chem 2017;65:2784–92. https://doi.org/10.1021/acs.jafc.6b04716.

[27] Packer NH, Lawson MA, Jardine DR, Redmond JW. A general approach to desalting oligosaccharides released from glycoproteins. Glycoconj J 1998;15:737–47. https://doi.org/10.1023/A:1006983125913.

[28] Robinson RC, Colet E, Tian T, Poulsen NA, Barile D. An improved method for the purification of milk oligosaccharides by graphitised carbon-solid phase extraction. Int Dairy J 2018;80:62–8. https://doi.org/10.1016/j.idairyj.2017.12.009.

[29] Perreault H, Costello CE. Stereochemical effects on the mass spectrometric behavior of native and derivatized trisaccharide isomers: comparisons with results from molecular modeling. J Mass

Spectrom 1999;34:184–97. https://doi.org/10.1002/(SICI)1096-9888(199903)34:3<184::AID-JMS783>3.0.CO;2-8.

[30] Yu RK, Koerner TAW, Ando S, Yohe HC, Prestegard JH. High-Resolution Proton NMR Studies of Gangliosides. J Biochem 1985;98:1367–73.

[31] Xu Y-F, Lu W, Rabinowitz JD. Avoiding Misannotation of In-Source Fragmentation Products as Cellular Metabolites in Liquid Chromatography–Mass Spectrometry-Based Metabolomics. Anal Chem 2015;87:2273–81. https://doi.org/10.1021/ac504118y.

[32] Hu C, Luo W, Xu J, Han X. Recognition and Avoidance of Ion Source-Generated Artifacts in Lipidomics Analysis. Mass Spectrom Rev 2022;41:15–31. https://doi.org/10.1002/mas.21659.

[33] Levin DS, Vouros P, Miller RA, Nazarov EG. Using a nanoelectrospray-differential mobility spectrometer-mass spectrometer system for the analysis of oligosaccharides with solvent selected control over ESI aggregate ion formation. J Am Soc Mass Spectrom 2007;18:502–11. https://doi.org/10.1016/j.jasms.2006.10.008.

[34] Puppala M, Ponder J, Suryanarayana P, Reddy GB, Petrash JM, LaBarbera DV. The Isolation and Characterization of β-Glucogallin as a Novel Aldose Reductase Inhibitor from Emblica officinalis. PLoS One 2012;7:e31399. https://doi.org/10.1371/journal.pone.0031399.

[35] Séveno M, Cabrera G, Triguero A, Burel C, Leprince J, Rihouey C, et al. Plant N-glycan profiling of minute amounts of material. Anal Biochem 2008;379:66–72. https://doi.org/10.1016/j.ab.2008.04.034.

[36] Antonio C, Larson T, Gilday A, Graham I, Bergström E, Thomas-Oates J. Hydrophilic interaction chromatography/electrospray mass spectrometry analysis of carbohydrate-related metabolites from Arabidopsis thaliana leaf tissue. Rapid Commun Mass Spectrom 2008;22:1399–407. https://doi.org/10.1002/rcm.3519.

[37] Liu M-H, Tong X, Wang J-X, Zou W, Cao H, Su W-W. Rapid separation and identification of multiple constituents in traditional Chinese medicine formula Shenqi Fuzheng Injection by ultra-fast liquid chromatography combined with quadrupole-time-of-flight mass spectrometry. J Pharm Biomed Anal 2013;74:141–55. https://doi.org/10.1016/j.jpba.2012.10.024.

[38] Jorge TF, Florêncio MH, Ribeiro-Barros AI, António C. Quantification and structural characterization of raffinose family oligosaccharides in Casuarina glauca plant tissues by porous

graphitic carbon electrospray quadrupole ion trap mass spectrometry. Int J Mass Spectrom 2017;413:127–34. https://doi.org/10.1016/j.ijms.2016.05.004.

[39] Ngoka LC, Gal JFrancois, Lebrilla CB. Effects of Cations and Charge Types on the Metastable Decay Rates of Oligosaccharides. Anal Chem 1994;66:692–8. https://doi.org/10.1021/ac00077a018.

[40] Cancilla MT, Penn SG, Carroll JA, Lebrilla CB. Coordination of Alkali Metals to Oligosaccharides Dictates Fragmentation Behavior in Matrix Assisted Laser Desorption Ionization/Fourier Transform Mass Spectrometry. J Am Chem Soc 1996;118:6736–45. https://doi.org/10.1021/ja9603766.

[41] Song T, Aldredge D, Lebrilla CB. A Method for In-Depth Structural Annotation of Human Serum Glycans That Yields Biological Variations. Anal Chem 2015;87:7754–62. https://doi.org/10.1021/acs.analchem.5b01340.

[42] Dong X, Zhou S, Mechref Y. LC-MS/MS Analysis of Permethylated Free Oligosaccharides and N-glycans Derived from Human, Bovine, and Goat Milk Samples. Electrophoresis 2016;37:1532–48. https://doi.org/10.1002/elps.201500561.

[43] Karlsson NG, Wilson NL, Wirth H-J, Dawes P, Joshi H, Packer NH. Negative ion graphitised carbon nano-liquid chromatography/mass spectrometry increases sensitivity for glycoprotein oligosaccharide analysis. Rapid Commun Mass Spectrom 2004;18:2282–92. https://doi.org/10.1002/rcm.1626.

[44] Ruhaak LR, Deelder AM, Wuhrer M. Oligosaccharide analysis by graphitized carbon liquid chromatography–mass spectrometry. Anal Bioanal Chem 2009;394:163–74. https://doi.org/10.1007/s00216-009-2664-5.

[45] Amicucci MJ, Nandita E, Galermo AG, Castillo JJ, Chen S, Park D, et al. A nonenzymatic method for cleaving polysaccharides to yield oligosaccharides for structural analysis. Nat Commun 2020;11:3963. https://doi.org/10.1038/s41467-020-17778-1.

[46] Cho BG, Peng W, Mechref Y. Separation of Permethylated O-Glycans, Free Oligosaccharides, and Glycosphingolipid-Glycans Using Porous Graphitized Carbon (PGC) Column. Metabolites 2020;10:433. https://doi.org/10.3390/metabo10110433.

[47] Sugimura N, Furuya A, Yatsu T, Igarashi Y, Aoyama R, Izutani C, et al. Observed adducts on positive mode direct analysis in real time mass spectrometry – Proton/ammonium adduct

selectivities of 600-sample in-house chemical library. Eur J Mass Spectrom (Chichester) 2017;23:4–10. https://doi.org/10.1177/1469066717693851.

Chapter III

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

(This chapter was published as a journal article "**Huang, Y.-P.**; Robinson, R.C.; Dias, F.F.G.; de Moura Bell, J.M.L.N.; Barile, D. *Foods* 2022, 11, 340, doi:10.3390/foods11030340.")

Abstract

Reverse-phase solid-phase extraction (SPE) is regularly used for separating and purifying food-derived oligosaccharides and peptides prior to liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis. However, the diversity in physicochemical properties of peptides may prevent the complete separation of the two types of analytes. Peptides present in the oligosaccharide fraction not only interfere with glycomics analysis but also escape peptidomics analysis. This work evaluated different SPE approaches for improving LC-MS/MS analysis of both oligosaccharides and peptides through testing on peptide standards and a food sample of commercial interest (proteolyzed almond extract). Compared with conventional reverse-phase SPE, mixed-mode SPE (reverse-phase/strong cation exchange) was more effective in retaining small/hydrophilic peptides and capturing them in the high-organic fraction and thus allowed the identification of more oligosaccharides and dipeptides in the proteolyzed almond extract, with satisfactory MS/MS confirmation. Overall, mixed-mode SPE emerged as the ideal method for simultaneously improving the identification of food-derived oligosaccharides and small peptides using LC-MS/MS analysis.

Keywords: peptidomics; glycomics; sample preparation; mixed-mode solid-phase extraction; LC-MS/MS; protein hydrolysates

3.1. Introduction

Oligosaccharides are carbohydrates consisting of 2 to 20 monosaccharide units and are widely found in plants and mammalian milk. These non-digestible carbohydrates have been studied for their prebiotic effect on the gut microbiota and their immunomodulatory effects [1,2].

Based on this potential, food products and supplements targeting human gut and digestive health are one of the fastest-growing segments in the food industry, with annual revenue of $39 billion in 2019, which is expected to increase to over $70 billion by 2027 [3]. Similarly, peptides are small fragments of proteins and are universally found in foods. Besides functioning as basic nutrients, peptides with specific structural features can also exhibit bioactivities. Peptides with beneficial activities, such as antimicrobial, antihypertensive, and anti-inflammatory, have been discovered in a wide range of foods [4–6].

Enzymatic hydrolysis is considered the preferred method in the food industry for increasing protein extraction yields, enhancing protein digestibility, reducing allergenicity, etc. [7–9]. Some peptides generated by enzymatic hydrolysis, have been shown to possess various bioactivities, such that antihypertensive and antibacterial peptides were identified in hypoallergenic infant formula, which had been partially or extensively hydrolyzed [10,11]. Therefore, hydrolysis techniques have been applied to several food products currently on the market [12,13].

The advancement of liquid chromatography-mass spectrometry (LC-MS) and automated data analysis enables the profiling of hundreds of peptides in a sample in only one run and is now widely used in bottom-up proteomics. To avoid ion suppression and ensure data quality in LC-MS analysis, appropriate sample preparation to eliminate interfering substances from complex food materials is indispensable and is regularly fulfilled with reversed-phase solid-phase extraction (SPE) [10,11]. Reverse-phase SPE can separate salts and low-molecular-weight carbohydrates (i.e., simple sugars and oligosaccharides) from peptides because only peptides are retained through the hydrophobic interaction. However, some small peptides, specifically di- and tripeptides, tend to pass through reverse-phase SPE with aqueous eluent and are not recovered in the final peptide eluate [14,15].

Peptide identification is most commonly conducted in the context of proteomics studies, which aim to profile the complete set of intact proteins in a sample, and their relative abundances. In bottom-up proteomics, usually only peptides comprising more than four amino acid residues are analyzed for identifying the originating proteins. Focusing the analysis on these longer peptide sequences is done for several practical purposes: since the proteolysis for proteomics is achieved using specific enzymes with well-defined cleavage sites, only limited amounts of smaller peptides are generated. Furthermore, the amino acid sequences of small peptides may be present in many proteins and lack uniqueness, so they are not suitable for verifying the presence of a particular protein. Finally, the algorithms used by MS-based proteomics software often cannot identify di- and tripeptides from tandem-MS data due to the relatively low number of fragment ions generated during fragmentation. In contrast, information about small peptides is significantly valuable for peptidomics, especially for the purpose of studying bioactive peptides. This interested originated from growing evidence showing that several small peptides exert bioactivities and may have a higher chance of surviving digestion as well as entering the blood circulation to exert bioactivity systemically [16–19]. Moreover, when food material is subject to enzymatic hydrolysis during food processing and then this is followed by the subsequent gastrointestinal digestion after ingestion, it can be expected that proteins will be extensively hydrolyzed and numerous small peptides will be generated. Therefore, small peptides should also be taken into consideration and be characterized when studying the bioactivity of proteolytic products.

In order to characterize bioactive peptides comprehensively using LC-MS, sample preparation approaches using reverse-phase SPE need to be modified for capturing shorter-length peptides [14,15]. One must also consider that foods often contain both peptides and oligosaccharides, such as milk and plant-based foods. Oligosaccharides can be naturally occurring,

generated during processing, or intentionally added as functional ingredients when the foods are lacking such compounds. For food products containing both oligosaccharides and abundant peptides, such as extensively hydrolyzed infant formula, the LC-MS analysis of oligosaccharides will be daunting due to the presence of interfering peptides. In fact, when the peptides are not completely separated from the oligosaccharides, they can cause ion suppression, impede oligosaccharide fragmentation in tandem MS analysis, and consequently diminish identification. A porous graphitized carbon (PGC) column is routinely used for chromatographic separation of oligosaccharides before and during LC-MS [20]. As some peptides strongly bind to PGC sorbent and are very difficult to elute, peptides can cause interferences and even decrease the binding capacity of the PGC column for oligosaccharides, in addition to potentially reducing the column life. Therefore, an effective fractionation of oligosaccharides and peptides would benefit the analysis of both types of analytes.

Incorporating specific binding mechanisms to assist the retention of small and hydrophilic peptides is a potential solution for a more effective fractionation of oligosaccharides and peptides. In theory, protonating peptides' carboxyl groups through acidification would allow most peptides to carry one or more net positive charge(s) and enable their interaction with cation exchange resins. Mixed-mode SPE, including retention mechanisms of both the reverse phase and strong cation exchange, was used for peptide enrichment prior to LC-MS analysis in a recent study, in which 25 peptides (including 4 tripeptides) were identified from Bifidobacterium cultures [21]. Peptide analysis using mixed-mode chromatography has also been reported in a few studies, although C18 reverse phase is still the most popular stationary phase [21–23]. However, its application towards the fractionation of oligosaccharides and peptides, especially small peptides, has yet to be evaluated. The objective of this study was to compare different SPE approaches, including mixed-

mode (reverse-phase/strong cation exchange) and conventionally used reverse-phase SPE, for their efficacy for fractionating peptides and oligosaccharides and therefore improving peptide and oligosaccharide LC-MS data quality.

3.2. Materials and methods

3.2.1. Materials

A peptide standard mixture (H2016; containing Gly-Tyr, leucine enkephalin (YGGFL), methionine enkephalin (YGGFM), and angiotensin II (DRVYIHPF)), oligosaccharide standards (raffinose pentahydrate and stachyose hydrate), invertase from baker's yeast (S. cerevisiae), trifluoroacetic acid (TFA), ammonia solution 25% (LC-MS LiChropur), ammonium formate (LC-MS LiChropur), and sodium acetate (molecular biology grade) were obtained from MilliporeSigma (St. Louis, MO, USA). Angiotensin I (DRVYIHPFHL) and neurotensin (pE-LYENKPRRPYIL) were obtained from GenScript (Piscataway, NJ, USA). Acetonitrile (ACN, Optima LC/MS grade), formic acid (Optima LC/MS grade), 50% (w/w) sodium hydroxide, and a Qubit protein assay kit were purchased from Thermo Fisher Scientific (Waltham, MA, USA). Melibiose and manninotriose were generated from raffinose and stachyose standards, respectively, with the treatment of invertase as described previously [24]. A proteolyzed almond extract was prepared from almond flour, at pilot-scale (~10 L of slurry), as described previously [25]. Briefly, almond flour was extracted with water, and "Neutral Protease 2 million" from Bacillus subtilis (BIO-CAT, Virginia, NY, USA), which randomly cleaves peptide bonds in protein structures, was added at an amount equal to 0.5% of the almond flour weight. The extraction was carried out in a 10 L jacketed glass reactor model CG-1965-610M (Chemglass Life Sciences LLC, Vineland, NJ, USA) with a 1:10 solids-to-liquid ratio at 50 °C and pH 9 and stirring at 120 rpm for 60 min. The slurry was separated into four fractions: the insoluble fraction, protein-rich fraction (protein

extract), cream, and free oil. The protein extract was used for examining SPE efficacy in this study and was named proteolyzed almond extract.

3.2.2. Comparison of procedures for protein removal

Ethanol precipitation and ultrafiltration were evaluated for their efficacy in removing proteins in the proteolyzed almond extract. For the precipitation method, 500 μL of the proteolyzed almond extract was mixed with 100 μL cold ethanol and incubated at 4 °C overnight. The mixture was then centrifuged at 4255× g at 4 °C for 30 min. The supernatant was separated and dried completely with a centrifugal evaporator at 30 °C and then dissolved with water. For the ultrafiltration method, 500 μL of the proteolyzed almond extract was either filtered directly with 3 kDa molecular weight cut-off (MWCO) centrifugal filter (Amicon, MilliporeSigma) or firstly filtered with 0.2 μm disk filter using a syringe and then filtered sequentially with 10 kDa and 3 kDa MWCO centrifugal filters (Amicon, MilliporeSigma). Centrifugal filtration was conducted at 13,000× g at 4 °C for 30 min.

3.2.3. Comparison of solid-phase extraction approaches

3.2.3.1. Reverse-phase solid-phase extraction

Three classic reverse-phase SPE cartridges, including Discovery DSC-18 with either 100 mg sorbent packed in a 1-mL tube (C18 100 mg; MilliporeSigma) or with 500 mg sorbent packed in a 3-mL tube (C18 500 mg; MilliporeSigma), Discovery DSC-8 with 100 mg sorbent packed in a 1-mL tube (C8 100 mg; MilliporeSigma), and a hydrophilic-lipophilic balanced SPE cartridge—Oasis HLB—with 60 mg sorbent packed in a 3-mL tube (HLB 60 mg; Waters, Milford, MA, USA), were tested with the procedures described in Supporting information Table 3.S1. Briefly, the cartridges were conditioned with pure ACN or ACN with either 0.1% TFA or 0.1% formic acid and then accordingly with water, 0.1% TFA in water, or 0.1% formic acid in water. Peptide

standard mixtures or supernatants of the proteolyzed almond extract prepared in water, 0.1% TFA, or 0.1% formic acid were loaded to the pre-conditioned SPE cartridges. The cartridges with loaded samples were flushed with three column volumes of water, 0.1% TFA in water, or 0.1% formic acid in water, and then accordingly with three column volumes of 80% ACN or 80% ACN containing either 0.1% TFA or 0.1% formic acid.

3.2.3.2. Mixed-Mode Solid-Phase Extraction

Three mixed-mode SPE cartridges comprising reverse-phase and strong cation exchange properties were tested using the procedures listed in the Supporting information Table 3.S2. These included Strata-X-C with 30 mg sorbent packed in a 1-mL tube (X-C 30 mg; Phenomenex, Torrance, CA, USA), Oasis MCX with 30 mg sorbent packed in a 1-mL tube (MCX 30 mg; Waters), and Discovery DSC-MCAX with 100 mg sorbent packed in a 1-mL tube (MCAX 100 mg; MilliporeSigma). For mixed-mode SPE, the cartridges were conditioned with ACN and then with either 0.1% TFA in water or 0.1% formic acid in water. Peptide samples prepared either in 0.1% TFA or 0.1% formic acid were loaded to the cartridges. The cartridges loaded with samples were firstly flushed with 3 mL of 0.1% TFA in water or 0.1% formic acid in water and then flushed with 3 mL of an eluent consisting of 40–50% ACN modified with either 1% ammonia or 250–375 mM ammonium formate. All fractions eluted from SPE cartridges were collected for analysis.

3.2.4. Analysis of peptide standards

Aqueous and high-organic fractions collected from reverse-phase or mixed-mode SPE were analyzed by either a Microflex LRF matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF; Bruker Daltonics, Billerica, MA, USA) mass spectrometer or an Agilent 6520 Accurate-Mass Q-TOF LC-MS with a Chip Cube interface (Agilent Technologies, Santa Clara, CA, USA) to determine the presence of peptide standards in each fraction. For the MALDI-TOF

MS analysis, 1 μL of the sample was mixed with 1 μL of α-cyano-4-hydroxycinnamic acid prepared in 30% ACN containing 0.07% TFA. The mixture (0.5 μL) was spotted on a ground steel target plate and dried under vacuum. The analysis was conducted with either linear mode or reflectron mode. Before analyzing the SPE fractions, the instrument was calibrated by the same peptide standard mixtures not subjected to SPE. For the LC-MS analysis, peptide standards were separated on an Agilent Zorbax 300SB-C18 capillary chip with a 40 nL trap (75 μm ×150 mm, 5 μm) at a flow rate of 300 nL min^{-1}. The mobile phase consisted of 3% ACN with 0.1% formic acid (v/v) (A) and 89.9% ACN with 0.1% formic acid (v/v) (B). The 40-min gradient with linear increase or decrease was programmed as follows: 0–2.3% B from 0.0–0.1 min; 2.3–15% B from 0.1–4.0 min; 15–22% B from 4.0–18.0 min; 22–60% B from 18.0–23.0 min; 60–100% B from 23.0–23.1 min; 100% B from 23.1–28 min; 100–0% B from 28.0–28.1 min; and 0% B from 28.1-40.0 min. Scan ranges were *m/z* 70–1800 at 8 spectra sec^{-1} for MS and from *m/z* 50–1800 at a precursor abundance dependent speed with a target of 25,000 count $spectrum^{-1}$ for MS/MS. Collision energy (CE; V) of (0.03 × (*m/z*) + 2) was used in tandem MS analysis for the top 10 ions in each cycle. The drying gas was set at 325 °C and 5 L min^{-1}. A capillary voltage of 1930 V was applied. Detection of peptide standards in the SPE fractions was determined by matching the retention times and the precursor *m/z* with the peptide standard mixtures not subjected to SPE.

3.2.5. Measuring the recovery of peptides

The efficacy of fractionating peptides and oligosaccharides by reverse-phase and mixed-mode SPE was evaluated with the breakthrough and recovery of peptides in the aqueous fraction and high-organic fraction, respectively, using the proteolyzed almond extract. Aqueous and high-organic fractions of the proteolyzed almond extract were dried with a centrifugal evaporator after collecting from SPE and redissolved with 50 μL of water. The peptide concentration in the

redissolved samples was measured by Qubit 4 Fluorometer (Thermo Fisher Scientific) using a Qubit Protein Assay Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. The breakthrough and recovery were calculated with the following formulas: peptide breakthrough = (total peptides in aqueous fraction/total peptides loaded to SPE) × 100%; peptide recovery = (total peptides in high-organic fraction/total peptides loaded to SPE) × 100%.

3.2.6. Measuring the recovery of oligosaccharides

Oligosaccharides in the aqueous fractions collected from reverse-phase or mixed-mode SPE cartridges using the proteolyzed almond extract were quantified for calculating the recovery of oligosaccharides. The aqueous fractions were directly analyzed after being brought to 5 or 10 mL in a volumetric flask for samples collected from 1 mL or 3 mL SPE cartridges, respectively. The quantification of two oligosaccharides, raffinose and stachyose, was carried out on a Thermo Fisher Dionex ICS-5000+ high-performance anion-exchange chromatography system with a CarboPac PA200 guard column (3 × 50 mm) and a CarboPac PA200 analytical column (3 × 250 mm). The mobile phase was composed of water (A), 200 mM sodium hydroxide (B), and 100 mM sodium hydroxide with 100 mM sodium acetate (C). The analytes were separated by isocratic elution at 25% B at a flow rate of 0.5 mL min^{-1} for 30 min. After the elution of oligosaccharides, the column was regenerated with a linear gradient from 25% B + 0% C to 50% B + 10% C in 5 min, followed by holding at 100% B for 5 min, and equilibrated with 25% B + 0% C for 10 min before the next injection. The oligosaccharides were quantified against calibration curves built with external standards (r^2 > 0.9995). The recovery was calculated by the formula (oligosaccharides in aqueous fraction/oligosaccharide in the original sample loaded to SPE) × 100%.

3.2.7. Characterization of oligosaccharides in the proteolyzed almond extract by LC-MS/MS

The aqueous fractions collected from reverse-phase and mixed-mode SPE containing oligosaccharides were further purified by non-porous graphitized carbon SPE (250 mg, 3-mL tube, Supelclean ENVI-Carb, MilliporeSigma). A non-porous graphitized carbon SPE cartridge was conditioned with 80% ACN, equilibrated with water, and then loaded with the aqueous fraction collected from reverse-phase and mixed-mode SPE. The non-porous graphitized carbon SPE cartridge was flushed with three column volumes of water to remove salts and acid. Oligosaccharides were eluted with two column volumes of 40% ACN, dried completely, and redissolved in water. The samples were analyzed with an Agilent 6520 Accurate-Mass Q-TOF LC-MS as described previously [26], with chromatographic separation at a flow rate of 300 nL min^{-1}. Oligosaccharides were characterized by examining the MS/MS fragments to determine their monosaccharide composition. Due to potential in-source fragmentation, extracted ion chromatographic peaks of oligosaccharides with various degrees of polymerization that possibly originated from the same oligosaccharide based on their monosaccharide compositions and co-eluted at the same retention time were considered as one identification [24]. Raffinose and stachyose were confirmed by comparing with the corresponding standards. Melibiose and manninotriose were confirmed by comparing with the disaccharide and the trisaccharide generated enzymatically from raffinose and stachyose standards, respectively, with treatment by invertase.

3.2.8. Characterization of peptides in the proteolyzed almond extract by LC-MS/MS

Peptide characterization for the proteolyzed almond extract was performed on an Agilent 6520 Accurate-Mass Q-TOF LC-MS. The peptide samples purified by reverse-phase or mixed-mode SPE were injected into an Agilent Zorbax 300SB-C18 chip (40 nL enrichment column, 75 μm × 150 mm; for comparing different protein removal approaches) or an Agilent Polaris-HR-

Chip (360 nL enrichment column, 75 μm × 150 mm; for comparing different SPE approaches) and separated by a mobile phase, consisting of 3% ACN with 0.1% formic acid (A) and 89.9% ACN with 0.1% formic acid (B), eluted at a flow rate of 300 nL min^{-1} with the following gradients: 0–2.3% B from 0–0.1 min; 2.3–8% B from 0.1–2.0 min; 8–37% B from 2.0–40.0 min; 37–48 % B from 40.0–45.0 min; 48–100% B from 45.0–45.1 min; 100% B from 45.1–50.0 min; 100–0% B from 50.0-50.1 min; and 0% B from 50.1–65.0 min. The scan range was *m/z* 70–1800 for MS and *m/z* 50–1800 for MS/MS. The scan speed was set at 8 spectra sec^{-1} for MS and varied with precursor abundance with a target of 25,000 count $spectrum^{-1}$ for MS/MS, respectively. The ESI source was operated on positive mode with a capillary voltage of 1950 V and drying gas at 325 °C and 5 L min^{-1}. The top 10 ions with the highest intensities in each cycle were selected for tandem MS analysis with the CE set by a formula of (CE (V) = 0.03 × (*m/z*) + 2).

3.2.9. Peptide data analysis

Peptide data analysis was performed with PEAKS Studio software (Bioinformatics Solutions Inc., Waterloo, ON, Canada). Medium-sized peptides, defined here as peptides with lengths ≥ 5 amino acid residues and below the upper limit that generally can be identified by LC-MS/MS (~50 amino acid residues), were identified through database search using the Uniprot database with the species name Prunus dulcis (both SwissProt and TrEMBL, accessed 6/20/2019). The mass error tolerance was 10 ppm and 0.02 Da for the precursor and fragment ions, respectively. The enzyme option was set as “None” along with an unspecific digestion mode. A maximum of two variable modifications, including oxidation (M), phosphorylation (STY), and deamidation (NQ), was allowed. The results were filtered with a false discovery rate of 1.0%.

Identification of dipeptides was achieved by de novo sequencing using PEAKS Studio followed by manual MS/MS spectral inspection. The settings for mass error tolerance, enzyme,

and digestion mode are the same as those used for database search. A maximum of one variable modification (oxidation (M), phosphorylation (STY), or deamidation (NQ)) was allowed.

3.2.10. Statistical analysis

Statistical analyses were performed in R (version 3.5.3). Single-factor analysis of variance and the subsequent pair-wise comparison with the Tukey method (significance level $\alpha = 0.05$) were conducted to compare the efficacy of different SPE approaches.

3.3. Results and discussion

3.3.1. Efficacy of different solid-phase extraction approaches in binding peptides

Although reverse-phase SPE is regularly used in peptide sample preparation, peptides comprising different types and numbers of amino acid residues may possess fairly distinct physicochemical properties (e.g., hydrophobicity and size) and therefore may not be completely recovered with reverse-phase SPE. To understand the retention capability of different SPE cartridges for various peptides, mixtures of peptide standards consisting of 2 to 13 amino acid residues were tested. Peptide standard mixtures were loaded to the pre-conditioned SPE cartridges, which were subsequently washed sequentially with an aqueous eluent and a high-organic eluent. Ideally, oligosaccharides and peptides should be present in the aqueous and high-organic fractions, respectively.

3.3.1.1. Reverse-Phase Solid-Phase Extraction

Table 3.1 presents the effect of acidic modifiers on the recovery of peptide standards from reverse-phase SPE; the aqueous and high-organic fractions were both modified with either 0.1% TFA or 0.1% formic acid. Leucine enkephalin and methionine enkephalin, two endogenous opioid peptide neurotransmitters, were only recovered in the high-organic fraction and not in the aqueous

fraction for all the three reverse-phase SPE tested (C18 100 mg, C18 500 mg, and HLB 60 mg), regardless of whether the eluents were modified with acid or not. However, when the eluent was not modified with acid, Gly-Tyr, a dipeptide exerting moderate inhibition against angiotensin-converting enzyme and dipeptidyl peptidase IV [27,28], was only detected in the aqueous fraction. TFA and formic acid increased the retention of Gly-Tyr on reverse-phase SPE, but only "C18 500 mg" flushed with 0.1% TFA in water led to its complete recovery in the high-organic fraction. As Gly-Tyr has a very small molecular size and lacks very hydrophobic side chains, it tended to pass through with aqueous eluent for reverse-phase SPE. When TFA was added to the eluent, the bulky negatively charged trifluoroacetate ions formed ion pairs with Gly-Tyr, which carried a positive charge on the N-terminal amine under a low pH environment, and, therefore, increased the retention of Gly-Tyr. HLB is a hydrophilic-lipophilic balanced copolymer, which provides slight hydrophilic interaction aside from reverse-phase retention and, therefore, was expected to help the retention of less hydrophobic small peptides. However, a complete recovery of Gly-Tyr in the high-organic fraction was not achieved with HLB regardless of the use of acidic modifiers, possibly due to the small size of Gly-Tyr and the relatively weak hydrophilicity of the sorbent.

Table 3.1. Detection of peptide standards in aqueous (aq) and high-organic (org) fractions collected from different solid-phase extraction techniques.

Solid-Phase Extraction	Glycyl Tyrosine		Leucine Enkephalin (YGGFL)		Methionine Enkephalin (YGGFM)		Angiotensin II (DRVYIHPF)	
	Aq	org	aq	org	aq	org	aq	org
no modifier								
C18 100 mg	✓[1]			✓		✓		✓(low)
C18 500 mg	✓			✓		✓		
HLB 60mg	✓			✓		✓		✓
FA[2] as modifier								
C18 100 mg	✓			✓		✓		✓
C18 500 mg	✓	✓		✓		✓		

HLB 60mg	✓	✓	✓	✓	✓
TFA as modifier					
C18 100 mg	✓	✓	✓	✓	✓
C18 500 mg		✓	✓	✓	✓
HLB 60 mg	✓	✓	✓	✓	✓
FA as modifier for aq; *NH_3 as modifier for org*					
X-C 30 mg		✓	✓	✓	✓

[1] Checkmark represents the detection of the peptide by MALDI-TOF MS or LC-QTOF MS from the corresponding fraction. [2] FA, formic acid; TFA, trifluoroacetic acid.

TFA also improved the recovery of angiotensin II, a vasoconstrictor hormone, in the high-organic fraction from C18 SPE. When the high-organic eluent was not modified with acid, angiotensin II was detected as a tiny peak in the high-organic fractions collected from "C18 100 mg" but not detected in the fraction collected from "C18 500 mg". The addition of either TFA or formic acid helped the elution of angiotensin II from "C18 100 mg", whereas only TFA enabled its elution from "C18 500 mg" SPE. In comparison, angiotensin II was successfully recovered by the high-organic eluent from HLB SPE cartridges, even when the eluent was not modified with acid. We hypothesize that residual silanol groups on the C18 sorbent led to the strong retention of angiotensin II, which carries two basic amino acid residues, arginine and histidine. A silanol group can release a proton and consequently carry a negative charge. The acidity of silanol groups varies with types of silanol and is influenced by other factors in sorbent manufacturing [29,30]. The ratio of deprotonated to protonated silanol groups should be higher at a neutral pH than at an acidic pH. Assuming the sorbents of "C18 500 mg" and "C18 100 mg" were identical, the residual silanol groups of "C18 500 mg" should be five times greater than "C18 100 mg". This could explain the stronger retention of Angiotensin II on the "C18 500 mg" than on the "C18 100 mg". The addition of acids reduced the deprotonated silanol groups and thus weakened the retention caused by residual silanol groups. Additionally, bulky trifluoroacetate ions further weakened the interaction

between deprotonated silanol groups and angiotensin II and therefore enabled the elution of angiotensin II from “C18 500 mg”. In contrast to C18 SPE, HLB SPE is packed with polymerized sorbents and has no silanol group, so the elution of angiotensin II was not affected by the acidic modifiers in the eluent.

3.3.1.2. Mixed-Mode Solid-Phase Extraction

Mixed-mode SPE has retention mechanisms of both reverse phase and strong cation exchange. In order to make peptides positively charged, samples must be acidified before being loaded onto the mixed-mode SPE cartridge to protonate both the N-terminal amino group and the C-terminal carboxyl group in peptides. On the other hand, to elute peptides from mixed-mode SPE, either basifying eluent (for deprotonating the N-terminal amino group) or increasing ionic strength of the eluent is necessary. When the peptide standards Gly-Tyr, leucine enkephalin, methionine enkephalin, angiotensin II, and angiotensin I (a precursor to angiotensin II) were applied to “X-C 30 mg”, all five peptides were not detected in the aqueous fraction flushed with an eluent of 0.1% formic acid in water and were exclusively recovered in the high-organic fraction flushed with an eluent containing 80% ACN and 1% ammonia (Tables 3.1 and 3.2). However, neurotensin, a regulatory peptide found in the central nervous system and the gastrointestinal tract, was not recovered in the high-organic fraction using an eluent containing 80% ACN and 1% ammonia (Table 3.2). We suggest that two arginine and one lysine residues of neurotensin restricted the elution of neurotensin from the mixed-mode SPE. With a pKa of 12.5, the side chain of arginine remained protonated during the elution using an eluent containing 80% ACN and 1% ammonia, which had a measured pH of 10.9. The net charge of neurotensin should be 2+ under these conditions, so neurotensin was still retained by the sulfonyl groups on the mixed-mode sorbent. Interestingly, although angiotensin II and angiotensin I also contain one arginine, the eluent

containing 1% ammonia was able to elute the two peptides. To deal with the strong retention of peptides containing multiple arginines, instead of flushing the mixed-mode SPE cartridge with basified eluent, flushing it with an eluent with increased ionic strength by adding ammonium formate as a modifier was tested. The fractions containing ammonium formate could not be analyzed by MALDI-TOF because the ionization of peptides was greatly suppressed by the salts. Instead, as ammonium formate is a volatile salt, the sample could be directly injected into LC-MS without further desalting. The results (Table 3.2) showed that 250 mM ammonium formate in 50% ACN was able to elute neurotensin from "MCAX 100 mg", whereas a higher ionic strength of the eluent (375 mM ammonium formate in 40% ACN) was required to elute neurotensin from "X-C 30 mg" and "MCX 30 mg". We decreased the concentration of ACN in the eluent to increase the ionic strength while avoiding the salt-induced liquid-liquid phase separation. Fortunately, 40–50% ACN still, at least partially, eluted all the peptides tested.

Table 3.2. Detection of peptide standards in high-organic fraction collected from mixed-mode solid-phase extraction[1] eluted by eluents modified by ammonia or ammonium formate.

Solid-Phase Extraction	Composition of High-Organic Eluent	Angiotensin I	Neurotensin
NH3 as modifier			
X-C 30 mg	80% ACN, 1% NH3	✓[2]	
MCX 30 mg	80% ACN, 1% NH3	✓	
MCAX 100 mg	80% ACN, 1% NH3	✓	
NH4COOH as modifier			
X-C 30 mg	50% ACN, 250 mM NH4COOH	✓	
MCX 30 mg	50% ACN, 250 mM NH4COOH		
MCAX 100 mg	50% ACN, 250 mM NH4COOH	✓	✓
X-C 30 mg	40% ACN, 375 mM NH4COOH	✓	✓
MCX 30 mg	40% ACN, 375 mM NH4COOH	✓	✓
MCAX 100 mg	40% ACN, 375 mM NH4COOH	✓	✓

[1] All the mixed-mode SPE were washed with 0.1% formic acid before eluting high-organic fraction. [2] Checkmark represents the detection of the peptide by MALDI-TOF MS or Q-TOF LC-MS from the corresponding fraction.

3.3.2. Evaluating oligosaccharide and peptide sample preparation approaches using the proteolyzed almond extract

3.3.2.1. Comparison of procedures for protein removal

When analyzing complex food samples, removing proteins before SPE avoids overloading the cartridges and the consequent ineffective binding of target analytes on SPE sorbents. We compared ethanol precipitation and ultrafiltration to evaluate their performance in protein removal using the proteolyzed almond extract. Although ultrafiltration is often used for fractionating peptides based on their sizes, we observed that a significant loss of peptides occurred when filtering the proteolyzed almond extract with a centrifugal filter (MWCO 3 kDa). The extracted ion chromatogram (EIC) peak areas of peptides in the filtrate were much lower than the ones treated by protein precipitation with ethanol (Supporting information Figure 3.S1). The differences in peak area between the samples from filtration and protein precipitation were directly related to the molecular weight of peptides. Even using a sequential filtration with 0.22 μm disk filter, 10 kDa, and 3 kDa, centrifugal filters did not prevent the loss of peptides, so it appeared that membrane fouling caused by insoluble particles and large molecules was not the main factor leading to the loss. Loss of opioid peptides with sizes well below the MWCO using centrifugal filters was also reported in a previous study [31]. The loss might be ascribed to peptide–peptide interaction and peptide aggregation due to the excessively high concentration of peptides on the membrane surface [32,33]. To avoid the risk of losing peptides at the step of protein removal, proteolyzed almond extract that underwent protein precipitation with ethanol was chosen for further studying the efficacy of different SPE approaches in improving the characterization of peptides and oligosaccharides.

3.3.2.2. Comparison of Solid-Phase Extraction Approaches for Improving Oligosaccharide Characterization

Oligosaccharides are very hydrophilic compounds due to the abundant hydroxyl groups in their molecular structures. Thus, oligosaccharides are not expected to be retained on hydrophobic SPE sorbents. For mixed-mode SPE, neutral and acidic oligosaccharides are generally uncharged at an acidic pH and are not retained by sulfonyl groups. It is worth noting that oligosaccharides that are positively charged under acidic pH, such as chitosan oligosaccharides, are expected to be retained by sulfonyl groups. Therefore, it is not suitable to use mixed-mode SPE for their purification. To evaluate the effectiveness in fractionating oligosaccharides and peptides with reverse-phase and mixed-mode SPE, we firstly measured the recovery of oligosaccharides from the proteolyzed almond extract. Raffinose and stachyose are two major oligosaccharides in almonds, and the standards are commercially available, so they were chosen for studying the recovery of oligosaccharides. The results showed that a complete or near-complete recovery of the two oligosaccharides was achieved for most SPE cartridges tested except for "HLB 60 mg" (Figure 3.1A,B). It is likely the cyclic amide providing hydrophilic interaction in the HLB sorbent slightly retained oligosaccharides and reduced their recovery.

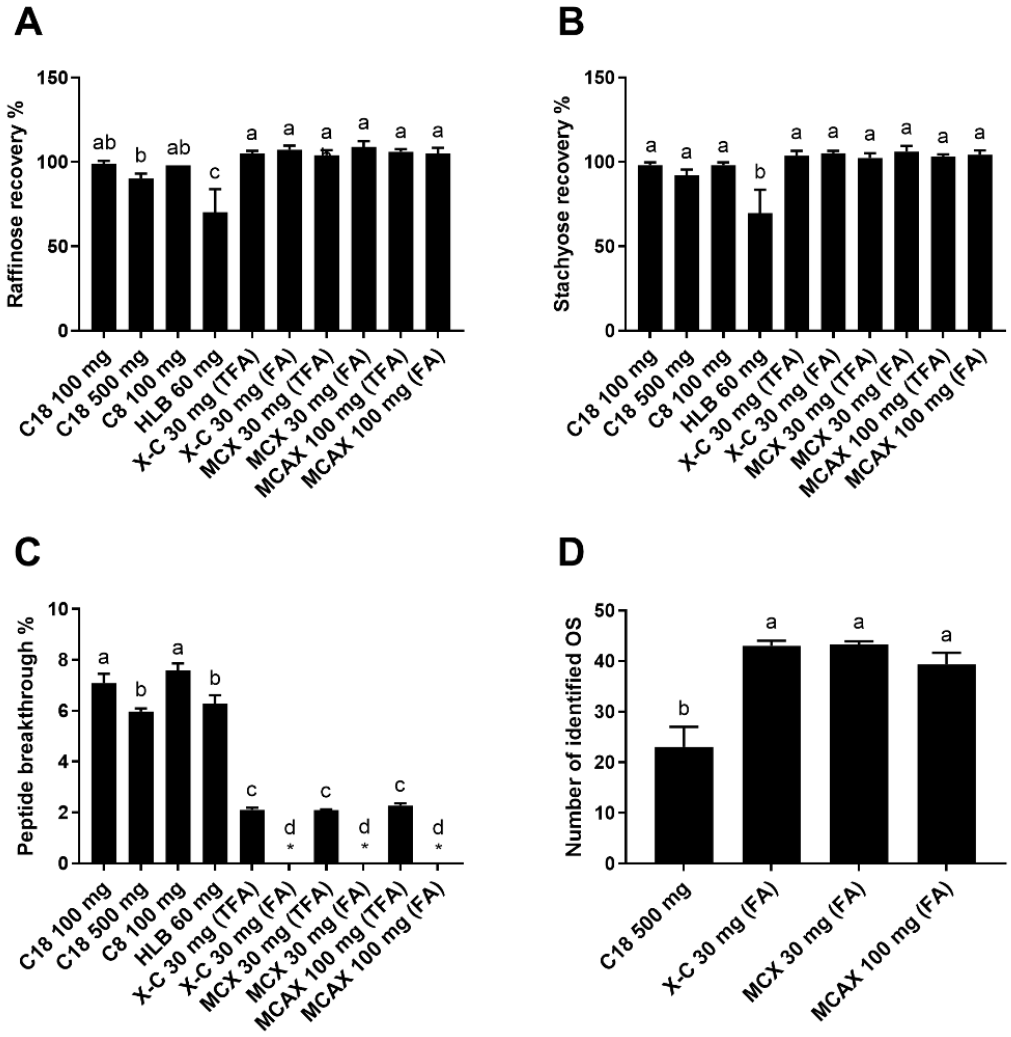


Figure 3.1. Recovery of raffinose (A) and stachyose (B), breakthrough of peptides (C), and number of oligosaccharides (OS) identified with MS/MS confirmation (D) of the aqueous fractions collected from different SPEs loaded with the proteolyzed almond extract. The asterisks indicate cases where the peptide breakthrough was lower than the detection limit (0.8%). Reverse-phase SPE (C18 100 mg, C18 500 mg, C8 100 mg, and HLB 60 mg) was conducted using TFA as a modifier. Mixed-mode SPE cartridges (X-C 30 mg, MCX 30 mg, and MCAX 30 mg) were eluted with either 0.1% TFA in water or 0.1% formic acid (FA) in water. Data are presented as mean ± standard deviation (n = 3). Different lowercase letters represent a significant difference at $p < 0.05$.

A low breakthrough of peptides in the fraction containing oligosaccharides is essential for effective analyses of both oligosaccharides and peptides. Not only would the peptide breakthrough interfere with the analysis of oligosaccharides but also the peptides would never be recovered in the high-organic fraction and therefore would escape characterization. The peptide breakthrough of reverse-phase SPE ranged from 6.0 to 7.6% (Figure 3.1C). Among the reverse-phase SPE cartridges, peptide breakthroughs of "C18 100 mg" and "C8 100 mg" were significantly higher than those of "C18 500 mg" and "HLB 60 mg". The lower peptide breakthrough of "C18 500 mg" than "C18 100 mg" indicated that a sufficient sorbent quantity could increase the retention of less hydrophobic peptides. It was also reported previously that the use of underloaded C18 SPE reduced the breakthrough of hydrophilic peptides [34]. The lower breakthrough of "HLB 60 mg" might arise from the higher binding capacity of the polymerized sorbent and the better retention of hydrophilic peptides of HLB sorbent than the silica-based sorbent [34]. Remarkably, the mixed-mode SPE resulted in a much lower peptide breakthrough, which was < 0.8% and 2.1–2.3% when eluting with 0.1% formic acid in water and 0.1% TFA in water, respectively (Figure 3.1C). The low breakthroughs of mixed-mode SPE indicated that the strong cation exchange property played an essential role in peptide retention. Contrary to C18 SPE, modification with 0.1% formic acid gave lower breakthroughs and seemed to be better for retaining small and hydrophilic peptides than with 0.1% TFA. This phenomenon might be caused by the competition between the sulfonyl groups on the mixed-mode sorbents and trifluoroacetate ions in the eluents. When TFA (pKa = 0.52) was used as a modifier, positively charged peptides could also form ion pairs with negatively charged trifluoroacetate ions, aside from being retained by sulfonyl groups. Hydrophobic interaction would therefore become the only retention mechanism for the ion pairs as the peptides' charge(s) was neutralized. Yet, the ion pairs of very small and hydrophilic peptides were still too

polar to be retained by hydrophobic interaction. In contrast, formic acid is a weaker acid (pKa = 3.75), so formic acid molecules in the eluent (pH ~2.6) were mostly undissociated and could not form ion pairs with peptides. Consequently, using TFA as a modifier resulted in higher peptide breakthroughs than using formic acid.

Overall, "C18 500 mg" cartridge gave a satisfactory oligosaccharide recovery and a relatively low peptide breakthrough compared with other reverse-phase SPE cartridges. Therefore, "C18 500 mg" was further compared with mixed-mode SPE on its capability in improving the data quality of oligosaccharide analysis with LC-MS. The oligosaccharide-containing fractions collected from "C18 500 mg" and mixed-mode SPE were subsequently purified with graphitized carbon SPE, a conventional step for oligosaccharide purification, and the oligosaccharides in both samples were analyzed with LC-MS. Several chromatographic peaks corresponding to oligosaccharides comprising hexoses and several peaks corresponding to released N-glycans were identified. We hypothesize that the N-glycans were released from glycopeptides during storage, possibly by glycoamidase originated from almonds as no glycoamidase was added to the proteolyzed almond extract. The reason for the presence of released N-glycans should be further investigated, but it is outside the scope of this study. However, regardless of the reason for the presence of released N-glycans, this diverse oligosaccharide composition is advantageous for our purpose of comparing different SPE approaches in the efficacy of improving oligosaccharide characterization.

A total of 44 oligosaccharides, including 19 oligosaccharides comprising hexoses and 25 oligosaccharides potentially being released N-glycans, were identified from the aqueous fractions by examining the tandem MS spectra to confirm the monosaccharide compositions (Supporting information Table 3.S3). The identified oligosaccharides may include some anomers, such as the

two anomers of melibiose. The number of oligosaccharides identified with tandem MS confirmation for the samples prepared with "X-C 30 mg" and "MCX 30 mg" was near twice the number obtained when using "C18 500 mg" (Figure 3.1D). This result indicates that the mixed-mode SPE may more effectively remove interferences and significantly improve the results of oligosaccharide analysis using LC-MS.

3.3.2.3. Comparison of Solid-Phase Extraction Approaches for Improving Peptide Characterization

3.3.2.3.1. Medium-Sized Peptides

After loading the deproteinized proteolyzed almond extract to the reverse-phase and mixed-mode SPE cartridges and eluting the fraction containing oligosaccharides, a high-organic solvent was used for peptides elution. The peptide recovery of reverse-phase SPE ranged between 77 and 84%, which was higher than that of "X-C 30 mg" eluted with 80% ACN and 1% ammonia (53% recovery) (Figure 3.2A). The lower recovery for the "X-C 30 mg" could be explained by the loss of peptides containing multiple arginine and lysine residues due to the strong retention by sulfonyl groups on the mixed-mode sorbent, as already observed with neurotensin. The comparison of the results for peptides identified by database search in the various samples prepared by employing different SPE approaches revealed that peptides containing multiple basic amino acids, such as arginine residues (e.g., peptide with sequence LDFVQPP**R**G**R**), and peptides containing one arginine along with multiple lysine residues (e.g., VTVP**K**EEE**KR**PQV**K**) were exclusively identified and detected in the samples prepared with reverse-phase SPE. Additionally, peptides containing multiple lysine residues, such as IMD**KI**K**E**K**K**LPGQH, were only partially recovered in the samples prepared with "X-C 30 mg" due to the strong retention by sulfonyl groups, as evidenced by the smaller peak areas than the reverse-phase SPE samples. In comparison,

peptides containing only one arginine or one lysine residue, such as LDFVQPPR, had comparable peak areas in the "X-C 30 mg" and all the reverse-phase SPE samples, which was in agreement with our prior results obtained by testing peptide standards.

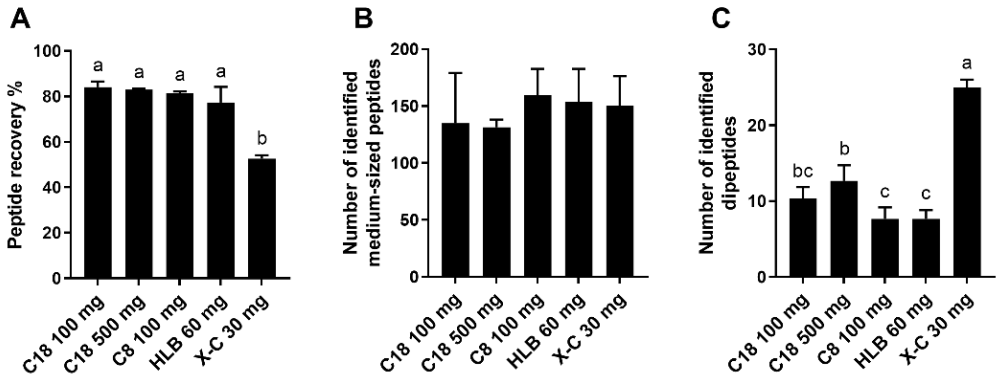


Figure 3.2. Recovery of proteins/peptides in high-organic fraction (A) and numbers of medium-sized peptides (peptide length ≥ 5 amino acid residues); (B) and dipeptides (C) identified by LC-MS from the proteolyzed almond extract collected from different SPE techniques. Reverse-phase SPE (C18 100 mg, C18 500 mg, C8 100 mg, and HLB 60 mg) was conducted using TFA as a modifier. Mixed-mode SPE (X-C-30 mg) was washed with 0.1% formic acid (aqueous fraction) and eluted with 80% ACN containing 1% ammonia for recovering peptides (high-organic fraction). Data are presented as mean ± standard deviation (n = 3). Different lowercase letters represent a significant difference at $p < 0.05$.

Despite the failure to identify some peptides containing multiple arginine and lysine residues in the "X-C 30 mg" high-organic fraction, due to the poor elution from the SPE, similar numbers of peptide sequences were overall obtained by database search (peptide length ≥ 5 amino acid residues) in the samples of "X-C 30 mg" and all the reverse-phase SPE methods tested (Figure 3.2B). This means that in reality peptides with numerous arginine and lysine residues likely accounted for an insignificant portion of all the identified peptides. However, due to the potential

of unsuccessful recovery of peptides containing multiple basic amino acid residues, one must consider the sample types and the analytes of interest when selecting SPE approaches. For example, mixed-mode SPE can be used for tryptic peptides that generally contain only one arginine or one lysine, as demonstrated in a previous study [21]. In contrast, mixed-mode SPE is not ideal when attempting to identify specifically cationic antimicrobial peptides, which typically contain several basic amino acid residues [35].

In an effort to improve the recovery of peptides containing multiple basic amino acid residues, we tried eluting peptides from "X-C 30 mg" using an eluent with increased ionic strength (40% ACN, 400 mM ammonium formate), but we were not able to measure the peptide recovery because ammonium formate caused a severe interference with the Qubit assay. Therefore, the samples eluted by 40% ACN and 400 mM ammonium formate were directly analyzed with LC-MS. Although the peptides containing two arginine residues were now successfully detected by LC-MS, ammonium formate in the final sample caused peak shape broadening and tailing for many peptides. Flushing the loading column with a larger volume of mobile phase during sample injection helped eliminate ammonium formate; however, as a consequence, the small and hydrophilic peptides were lost, thus eliminating a major goal of using mixed-mode SPE.

3.3.2.3.2. Small Peptides

Peptides with four or fewer amino acid residues can be identified from mass spectral data by de novo identification. In the present study, we focused on dipeptides because it is generally more challenging to retain them by SPE with hydrophobic interactions than tri- and tetrapeptides. We were able to identify 30 dipeptides in the proteolyzed almond extract. All the 30 dipeptides were identified in the samples prepared with "X-C 30 mg", eluted with 80% ACN and 1% NH3, with satisfying MS/MS confirmation, whereas the numbers were substantially lower for the

samples prepared with reverse-phase SPE (9–17 dipeptides) (Table 3.3 and Figure 3.2C). Several of the dipeptides exclusively identified in the samples prepared by "X-C 30 mg" were early-eluting peptides comprised of hydrophilic amino acid residues, such as glutamine. The improved retention of hydrophilic dipeptides achieved with "X-C 30 mg" can be attributed to the strong cation exchange property of the sorbent. The analysis of small peptides has recently been recognized as a major challenge in MS-based analysis [36,37]. Even though this is due to the convergence of many factors (including sample preparation, peptide enrichment, and data analysis), recovering small peptides using an appropriate sample preparation strategy is one of the critical steps that can help overcome such hurdle. This study proposed the feasibility of using mixed-mode SPE as a strategy to recover small peptides and demonstrated its success in enabling further analysis of small peptides by LC-MS.

Table 3.3. Dipeptides identified with LC-MS/MS in the high-organic fractions, of the proteolyzed almond extract, prepared with each solid-phase extraction technique.[1]

Peptide Sequence	Retention Time (min)	C18 100 mg	C18 500 mg	C8 100 mg	HLB 60 mg	X-C 30 mg
Gln-Gln	2.00					✓[2]
Gly-Gln	2.00					✓
Ala-Pro	2.21					✓
Gly-Val	2.28					✓
Lxx-Glu[3]	3.24					✓
Ser-Tyr	3.58		✓			✓
Gly-Tyr	3.65		✓			✓
Val-Pro	3.84		✓			✓
Thr-Tyr	3.93					✓
Ser-Lxx	4.80					✓
Gly-Lxx	5.13		✓			✓
Ala-Lxx	5.23					✓
Thr-Lxx	6.17		✓			✓
Val-Tyr	6.28	✓	✓			✓
Lxx-Val	6.83	✓				✓
Ser-Phe	6.91					✓
Gly-Phe	7.12	✓	✓		✓	✓

Ala-Phe	7.27					✓
Lxx-Pro	7.38	✓	✓			✓
Val-Lxx	7.66		✓			✓
Phe-Pro	9.45	✓	✓	✓	✓	✓
Lxx-Phe	10.42	✓		✓		✓
Trp-Pro	11.08	✓	✓	✓	✓	✓
Tyr-Trp	11.72					✓
Val-Met	11.74	✓	✓	✓	✓	✓
Lxx-Phe	11.88	✓	✓	✓	✓	✓
Lxx-Trp	12.75	✓	✓	✓	✓	✓
Lxx-Trp	13.39	✓	✓	✓	✓	✓
Phe-Phe	13.40	✓	✓	✓	✓	✓
Phe-Trp	15.06	✓	✓	✓	✓	✓

[1] Reverse-phase SPE (C18 100 mg, C18 500 mg, C8 100 mg, and HLB 60 mg) was eluted with 0.1% TFA (aqueous fraction) followed by 80% ACN modified with 0.1% TFA (high-organic fraction). Mixed-mode SPE (X-C 30 mg) was eluted with 0.1% formic acid (aqueous fraction), and 80% ACN was modified with 1% NH3 (high-organic fraction).

[2] Checkmarks represent peptide identification, with MS/MS confirmation in at least one of the three replicates.

[3] Lxx denotes that the amino acid residue could be either Leu or Ile.

3.4. Conclusions

Disorders caused by dysregulated gastrointestinal microbiomes are increasingly common. Currently, such disorders are treated by small-molecule antimicrobial drugs, which unfortunately lack selectivity, killing both pathogenic and commensal organisms and thus leading to further disruption of the microbiome. Therefore, there is growing interest in modulating gut health with novel food products rich in functional ingredients such as peptides and oligosaccharides, which have significant potential to impact human health. In order to comprehensively characterize small and medium-sized bioactive peptides and oligosaccharides in foods using LC-MS, sample preparation approaches using various SPE need to be adapted for capturing all compounds of interest. The proteolyzed almond extract was selected as a model because almond proteins contain a high proportion of hydrophilic amino acids, resulting in a more difficult peptide recovery via

conventional reverse-phase SPE. Therefore, having established a successful model system on a challenging food material, these findings could be universally applied to other abundant matrices such as animal proteins, which are known to contain more hydrophobic amino acid residues. Based on the evaluation in this study, when studying proteolyzed food samples, in which oligosaccharides would be found together with abundant peptides, mixed-mode SPE should be preferred over reverse-phase SPE because it led to lower peptide breakthrough and therefore improved oligosaccharide identification validated by tandem MS confirmation. When the purpose of characterization is mainly focused on the discovery of bioactive peptides, factors such as peptide size, hydrophobicity, and charge must be taken into account. For peptides with sufficient hydrophobicity (which are generally medium-sized peptides), C18 SPE with an adequate amount of sorbent leads to more robust results. Although mixed-mode SPE could render a similar number of medium-sized peptides as reverse-phase SPE, failure to identify peptides containing multiple basic amino acid residues might be a concern. Nevertheless, when small and hydrophilic peptides are of interest, mixed-mode SPE remains the ideal choice because of its effective retention of these types of peptides. In summary, this study compared the efficacy of separating oligosaccharides and peptides with reverse-phase and mixed-mode SPE approaches, providing a useful guide for selecting specific sorbents and solvents based on the properties of food samples and compounds of interest.

Author contributions

Conceptualization, Y.-P.H. and D.B.; methodology, Y.-P.H., R.C.R. and D.B.; formal analysis, Y.-P.H.; investigation, Y.-P.H., R.C.R. and F.F.G.D.; resources, D.B.; data curation, Y.-P.H.; writing—original draft preparation, Y.-P.H.; writing—review and editing, R.C.R., F.F.G.D.,

J.M.L.N.d.M.B. and D.B.; visualization, Y.-P.H.; supervision, D.B.; project administration, J.M.L.N.d.M.B. and D.B.; funding acquisition, J.M.L.N.d.M.B. and D.B. All authors have read and agreed to the published version of the manuscript.

Funding

Funding for this study was made possible by the U.S. Department of Agriculture's (USDA) Agricultural Marketing Service through grant AM170100XXXXG011 and USDA NIFA project CA−DFST-2187-H. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the USDA.

Data availability statement

The data presented in this study are available in article and supplementary material.

Conflicts of interest

The authors declare no conflict of interest.

Supplementary materials

Table 3.S1. Procedures of different reverse-phase solid-phase extraction strategies.

solid-phase extraction step	no modifier				TFA as modifier				FA as modifier			
	C18 100 mg	C18 500 mg	C8 100 mg	HLB 60 mg	C18 100 mg	C18 500 mg	C8 100 mg	HLB 60 mg	C18 100 mg	C18 500 mg	C8 100 mg	HLB 60 mg
condition	3 column volumes of ACN				3 column volumes of ACN with 0.1 % TFA				3 column volumes of ACN with 0.1 % FA			
equilibration	3 column volumes of water				3 column volumes of water with 0.1 % TFA				3 column volumes of water with 0.1 % FA			
sample loading	samples prepared in water				samples prepared in 0.1% TFA				samples prepared in 0.1% FA			
oligosaccharide elution (aq)	3 column volumes of water				3 column volumes of 0.1% TFA				3 column volumes of 0.1% FA			
peptide elution (org)	3 column volumes of 80% ACN				3 column volumes of 80% ACN with 0.1% TFA				3 column volumes of 80% ACN with 0.1% FA			

[1] TFA, trifluoroacetic acid; FA, formic acid; ACN, acetonitrile; aq, aqueous fraction; org, high-organic fraction.

Table 3.S2. Procedures of different mixed-mode solid-phase extraction strategies.

solid-phase extraction step	FA and NH_3 as modifiers			TFA and NH_3 as modifiers			FA and NH_4COOH as modifiers		
	X-C 30 mg	MCX 30 mg	MCAX 100 mg	X-C 30 mg	MCX 30 mg	MCAX 100 mg	X-C 30 mg	MCX 30 mg	MCAX 100 mg
condition	3 column volumes of ACN			3 column volumes of ACN			3 column volumes of ACN		
equilibration	3 column volumes of water with 0.1% FA			3 column volumes of water with 0.1% TFA			3 column volumes of water with 0.1% FA		
sample loading	samples prepared in 0.1% FA			samples prepared in 0.1% TFA			samples prepared in 0.1% FA		
oligosaccharide elution (aq)	3 column volumes of 0.1% FA			3 column volumes of 0.1% TFA			3 column volumes of 0.1% FA		
peptide elution (org)	3 column volumes of 80% ACN with 1% NH_3			3 column volumes of 80% ACN with 1% NH_3			3 column volumes of 50% ACN with 250 mM NH_4COOH or 40% ACN with 375 mM NH_4COOH		

[1] FA, formic acid; TFA, trifluoroacetic acid; ACN, acetonitrile; aq, aqueous fraction; org, high-organic fraction.

Table 3.S3. List of oligosaccharides identified from proteolyzed aqueous almond extract with LC-MS

monosaccharide composition[1]	retention time (min)	observed MS[1] ions
Hex_2 (melibiose)[2]	0.7	325, 685, 343, 667, 487, 649, 505
Hex_2 (melibiose)	1.2	325, 685, 343, 667, 487, 505, 649
Hex_3 (manninotriose)	3.2	325, 487
Hex_3	4.0	505, 343
Hex_3 (manninotriose)	4.6	325, 487
$Hex_2HexNAc_2Fuc_1$	8.0	895
Hex_4	8.5	667
Hex_3 (raffinose)	9.4	325, 343, 685, 667, 505, 487
Hex_3	10.7	505, 829
$Hex_3HexNAc_4Fuc_1Xyl_1$	10.8	798
$Hex_4HexNAc_2Fuc_1Xyl_1$	10.8	1351, 676
$Hex_3HexNAc_2Fuc_1Xyl_1$	11.0	1189, 1043
Hex_4 (stachyose)	11.3	325, 487, 505, 343
$Hex_4HexNAc_4Fuc_2Xyl_1$	12.4	952
$Hex_4HexNAc_4Fuc_1Xyl_1$	12.4	879
$Hex_3HexNAc_3Fuc_1Xyl_1$	12.5	696
$Hex_4HexNAc_4Fuc_1Xyl_1$	13.1	879
$Hex_4HexNAc_4Fuc_2Xyl_1$	13.2	952
$Hex_2HexNAc_2Fuc_1Xyl_1$	13.3	1027
$Hex_3HexNAc_4Fuc_1Xyl_1$	13.7	798
Hex_4	14.1	649, 667, 487, 325, 505
$Hex_5HexNAc_4Fuc_3Xyl_1$	14.5	1106
$Hex_3HexNAc_2Fuc_1Xyl_1$	14.7	1189, 505, 487, 829
$Hex_2HexNAc_2Xyl_1$	14.9	881
$Hex_3HexNAc_4Xyl_1$	15.0	725
Hex_4	15.2	649, 667
Hex_4	15.5	649, 487, 325
$Hex_3HexNAc_4$	15.9	659
$Hex_3HexNAc_4Xyl_1$	16.3	725
Hex_4	16.4	505, 325, 487, 343, 667
$Hex_3HexNAc_2Xyl_1$	16.4	1043
$Hex_3HexNAc_2$	16.6	911
$Hex_2HexNAc_2Xyl_1$	16.7	881
$Hex_5HexNAc_4Fuc_3Xyl_1$	17.4	1106
$Hex_3HexNAc_2Xyl_1$	18.3	1043
$Hex_3HexNAc_2$	18.4	911
$Hex_3HexNAc_2Xyl_1$	18.9	1043
$Hex_5HexNAc_4Fuc_2Xyl_1$	20.1	1033
Hex_4	25.8	667, 649, 325, 487
Hex_5	26.0	829

Hex_4	26.1	667, 649, 487
Hex_4	26.4	667, 649, 487
Hex_5	26.6	829
Hex_5	30.6	829, 649

[1] Abbreviation for monosaccharides: Hex, hexose; HexNAc, N-acetylhexosamine; Fuc: fucose; Xyl: xylose.
[2] Melibiose and manninotriose were confirmed by comparing with melibiose and manninotriose generated enzymatically from raffinose and stachyose standards, respectively, with treatment by invertase. Raffinose and stachyose were confirmed by comparing with the corresponding standards.

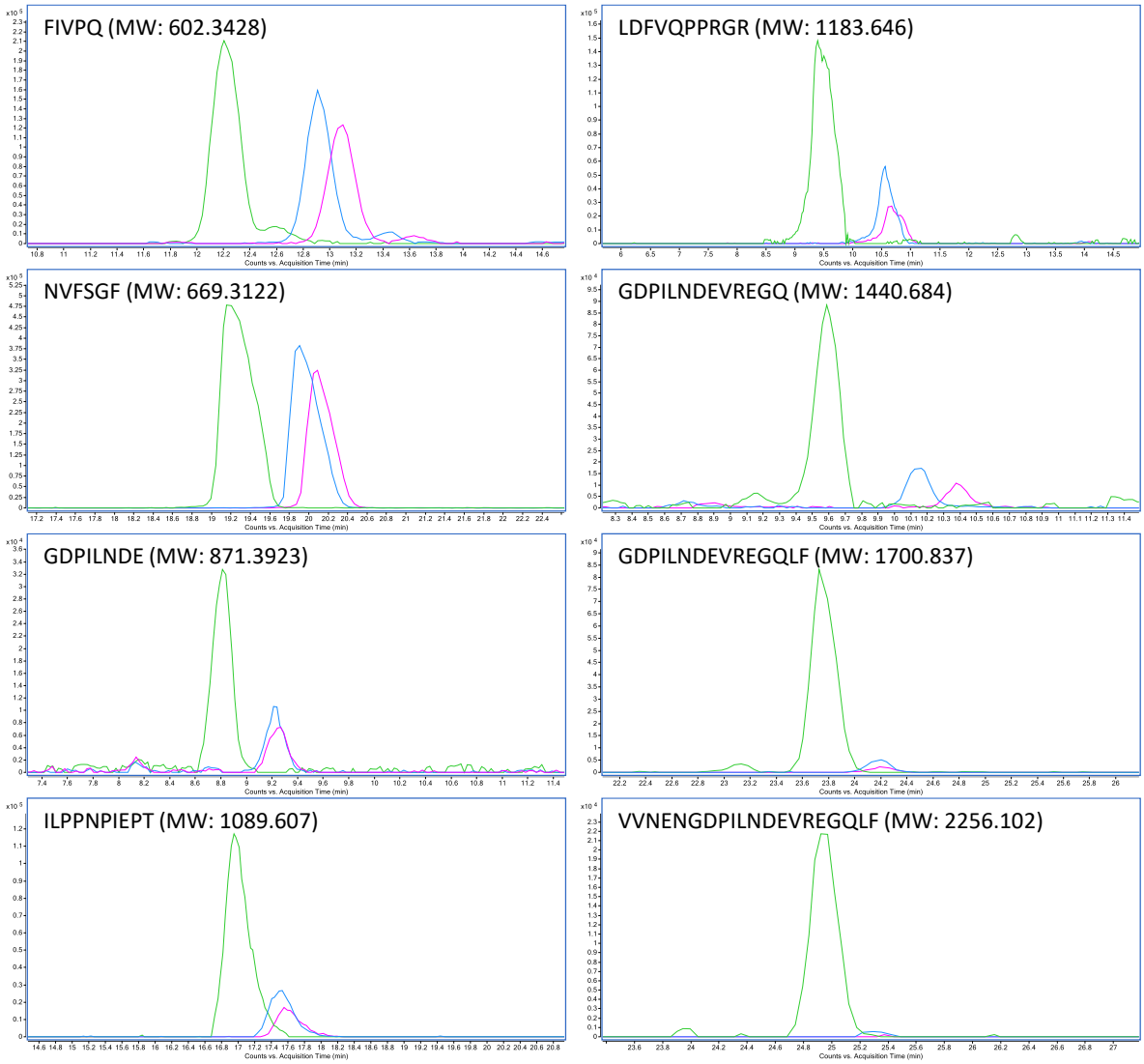


Figure 3.S1. Extracted ion chromatograms (EIC) of selected peptides from the proteolyzed aqueous almond extract purified with different protein removal strategies (green line: protein precipitation; blue line: centrifugal filtration (MWCO 3,000 Da); pink line: sequential filtration with disk filter (0.22 μm) and centrifugal filter (MWCO 10,000 Da and then MWCO 3,000 Da)) followed by C18 500 mg SPE (TFA as modifier).

References

1. Chichlowski, M.; German, J.B.; Lebrilla, C.B.; Mills, D.A. The Influence of Milk Oligosaccharides on Microbiota of Infants: Opportunities for Formulas. *Annu. Rev. Food Sci. Technol.* **2011**, *2*, 331–351. https://doi.org/10.1146/annurev-food-022510-133743.

2. Mussatto, S.I.; Mancilha, I.M. Non-Digestible Oligosaccharides: A Review. *Carbohydr. Polym.* **2007**, *68*, 587–597. https://doi.org/10.1016/j.carbpol.2006.12.011.

3. Human Milk Oligosaccharides Market Size, Share & Trends Analysis Report. Available online: https://www.grandviewresearch.com/industry-analysis/human-milk-oligosaccharides-market (accessed on 6 November 2021).

4. Miralles, B.; Amigo, L.; Recio, I. Critical Review and Perspectives on Food-Derived Antihypertensive Peptides. *J. Agric. Food Chem.* **2018**, *66*, 9384–9390. https://doi.org/10.1021/acs.jafc.8b02603.

5. Guha, S.; Majumder, K. Structural-Features of Food-Derived Bioactive Peptides with Anti-Inflammatory Activity: A Brief Review. *J. Food Biochem.* **2019**, *43*, e12531. https://doi.org/10.1111/jfbc.12531.

6. Phazang, P.; Negi, N.P.; Raina, M.; Kumar, D. Plant Antimicrobial Peptides: Next-Generation Bioactive Molecules for Plant Protection. In *Phyto-Microbiome in Stress Regulation*; Kumar, M., Kumar, V., Prasad, R., Eds.; Environmental and Microbial Biotechnology; Springer: Singapore, 2020; pp. 281–293. https://doi.org/10.1007/978-981-15-2576-6_14.

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

8. Koopman, R.; Crombach, N.; Gijsen, A.P.; Walrand, S.; Fauquant, J.; Kies, A.K.; Lemosquet, S.; Saris, W.H.; Boirie, Y.; van Loon, L.J. Ingestion of a Protein Hydrolysate Is Accompanied by an Accelerated in Vivo Digestion and Absorption Rate When Compared with Its Intact Protein. *Am. J. Clin. Nutr.* **2009**, *90*, 106–115. https://doi.org/10.3945/ajcn.2009.27474.

9. Bahna, S.L. Hypoallergenic Formulas: Optimal Choices for Treatment versus Prevention. *Ann. Allergy Asthma. Immunol.* **2008**, *101*, 453–459. https://doi.org/10.1016/S1081-120660281-5.

10. Català-Clariana, S.; Benavente, F.; Giménez, E.; Barbosa, J.; Sanz-Nebot, V. Identification of Bioactive Peptides in Hypoallergenic Infant Milk Formulas by Capillary Electrophoresis–Mass Spectrometry. *Anal. Chim. Acta* **2010**, *683*, 119–125. https://doi.org/10.1016/j.aca.2010.10.002.

11. Wada, Y.; Lönnerdal, B. Bioactive Peptides Released by in Vitro Digestion of Standard and Hydrolyzed Infant Formulas. *Peptides* **2015**, *73*, 101–105. https://doi.org/10.1016/j.peptides.2015.09.005.

12. Aluko, R.E. 15—Food Protein-Derived Peptides: Production, Isolation, and Purification. In *Proteins in Food Processing*, 2nd ed.; Yada, R.Y., Ed.; Woodhead Publishing Series in Food Science, Technology and Nutrition; Woodhead Publishing: Cambridge, UK. 2018; pp. 389–412. https://doi.org/10.1016/B978-0-08-100722-8.00016-4.

13. Lafarga, T.; Hayes, M. Bioactive Protein Hydrolysates in the Functional Food Ingredient Industry: Overcoming Current Challenges. *Food Rev. Int.* **2017**, *33*, 217–246. https://doi.org/10.1080/87559129.2016.1175013.

14. Piovesana, S.; Capriotti, A.L.; Cerrato, A.; Crescenzi, C.; La Barbera, G.; Laganà, A.; Montone, C.M.; Cavaliere, C. Graphitized Carbon Black Enrichment and UHPLC-MS/MS Allow to Meet the Challenge of Small Chain Peptidomics in Urine. *Anal. Chem.* **2019**, *91*, 11474–11481. https://doi.org/10.1021/acs.analchem.9b03034.

15. Piovesana, S.; Cerrato, A.; Antonelli, M.; Benedetti, B.; Capriotti, A.L.; Cavaliere, C.; Montone, C.M.; Laganà, A. A Clean-up Strategy for Identification of Circulating Endogenous Short Peptides in Human Plasma by Zwitterionic Hydrophilic Liquid Chromatography and Untargeted Peptidomics Identification. *J. Chromatogr. A* **2020**, *1613*, 460699. https://doi.org/10.1016/j.chroma.2019.460699.

16. Matsui, T.; Tamaya, K.; Seki, E.; Osajima, K.; Matsumoto, K.; Kawasaki, T. Val-Tyr As A Natural Antihypertensive Dipeptide Can Be Absorbed Into The Human Circulatory Blood System. *Clin. Exp. Pharmacol. Physiol.* **2002**, *29*, 204–208. https://doi.org/10.1046/j.1440-1681.2002.03628.x.

17. Ruiz, J.Á.G.; Ramos, M.; Recio, I. Angiotensin Converting Enzyme-Inhibitory Activity of Peptides Isolated from Manchego Cheese. Stability under Simulated Gastrointestinal Digestion. *Int. Dairy J.* **2004**, *14*, 1075–1080. https://doi.org/10.1016/j.idairyj.2004.04.007.

18. Ohsawa, K.; Satsu, H.; Ohki, K.; Enjoh, M.; Takano, T.; Shimizu, M. Producibility and Digestibility of Antihypertensive β-Casein Tripeptides, Val-Pro-Pro and Ile-Pro-Pro, in the Gastrointestinal Tract: Analyses Using an in Vitro Model of Mammalian Gastrointestinal Digestion. *J. Agric. Food Chem.* **2008**, *56*, 854–858. https://doi.org/10.1021/jf072671n.

19. Sato, K. Structure, Content, and Bioactivity of Food-Derived Peptides in the Body. *J. Agric. Food Chem.* **2018**, *66*, 3082–3085. https://doi.org/10.1021/acs.jafc.8b00390.

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

21. Chen, S.; Huang, G.; Liao, W.; Gong, S.; Xiao, J.; Bai, J.; Hsiao, W.L.W.; Li, N.; Wu, J.L. Discovery of the Bioactive Peptides Secreted by Bifidobacterium Using Integrated MCX Coupled with LC–MS and Feature-Based Molecular Networking. *Food Chem.* **2021**, *347*, 129008. https://doi.org/10.1016/j.foodchem.2021.129008.

22. Gilar, M.; Yu, Y.Q.; Ahn, J.; Fournier, J.; Gebler, J.C. Mixed-Mode Chromatography for Fractionation of Peptides, Phosphopeptides, and Sialylated Glycopeptides. *J. Chromatogr. A* **2008**, *1191*, 162–170. https://doi.org/10.1016/j.chroma.2008.01.061.

23. Li, X.; Guo, Z.; Sheng, Q.; Xue, X.; Liang, X. Sequential Elution of Multiply and Singly Phosphorylated Peptides with Polar-Copolymerized Mixed-Mode RP18/SCX Material. *Analyst* **2012**, *137*, 2774–2776. https://doi.org/10.1039/C2AN35247H.

24. Huang, Y.P.; Robinson, R.C.; Barile, D. Food Glycomics: Dealing with Unexpected Degradation of Oligosaccharides during Sample Preparation and Analysis. *J. Food Drug Anal.* **2022**, *30*, 62–76, https://doi.org/10.38212/2224-6614.3393.

25. de Almeida, N.M.; Dias, F.F.G.; Rodrigues, M.I.; de Moura Bell, J.M.L.N. Effects of Processing Conditions on the Simultaneous Extraction and Distribution of Oil and Protein from Almond Flour. *Processes* **2019**, *7*, 844. https://doi.org/10.3390/pr7110844.

26. Sischo, W.M.; Short, D.M.; Geissler, M.; Bunyatratchata, A.; Barile, D. Comparative Composition, Diversity, and Abundance of Oligosaccharides in Early Lactation Milk from Commercial Dairy and Beef Cows. *Int. J. Dairy Sci.* **2017**, *100*, 3883–3892. https://doi.org/10.3168/jds.2016-12388.

27. Cheung, H.S.; Wang, F.L.; Ondetti, M.A.; Sabo, E.F.; Cushman, D.W. Binding of Peptide Substrates and Inhibitors of Angiotensin-Converting Enzyme. Importance of the COOH-Terminal Dipeptide Sequence. *J. Biol. Chem.* **1980**, *255*, 401–407.

28. Lan, V.T.T.; Ito, K.; Ohno, M.; Motoyama, T.; Ito, S.; Kawarasaki, Y. Analyzing a Dipeptide Library to Identify Human Dipeptidyl Peptidase IV Inhibitor. *Food Chem.* **2015**, *175*, 66–73. https://doi.org/10.1016/j.foodchem.2014.11.131.

29. Méndez, A.; Bosch, E.; Rosés, M.; Neue, U.D. Comparison of the Acidity of Residual Silanol Groups in Several Liquid Chromatography Columns. *J. Chromatogr. A* **2003**, *986*, 33–44. https://doi.org/10.1016/S0021-967301899-X.

30. Sulpizi, M.; Gaigeot, M.P.; Sprik, M. The Silica–Water Interface: How the Silanols Determine the Surface Acidity and Modulate the Water Properties. *J. Chem. Theory Comput.* **2012**, *8*, 1037–1047. https://doi.org/10.1021/ct2007154.

31. Benavente, F.; Medina-Casanellas, S.; Barbosa, J.; Sanz-Nebot, V. Investigation of Commercial Sorbents for the Analysis of Opioid Peptides in Human Plasma by On-Line SPE-CE. *J. Sep. Sci.* **2010**, *33*, 1294–1304. https://doi.org/10.1002/jssc.200900669.

32. Arakawa, T.; Ejima, D.; Akuta, T. Protein Aggregation under High Concentration/Density State during Chromatographic and Ultrafiltration Processes. *Int. J. Biol. Macromol.* **2017**, *95*, 1153–1158. https://doi.org/10.1016/j.ijbiomac.2016.11.005.

33. Groleau, P.E.; Lapointe, J.F.; Gauthier, S.F.; Pouliot, Y. Effect of Aggregating Peptides on the Fractionation of β-LG Tryptic Hydrolysate by Nanofiltration Membrane. *J. Membr. Sci.* **2004**, *234*, 121–129. https://doi.org/10.1016/j.memsci.2004.01.016.

34. Guo, X.; Kristal, B.S. The Use of Underloaded C18 Solid-Phase Extraction Plates Increases Reproducibility of Analysis of Tryptic Peptides from Unfractionated Human Plasma. *Anal. Biochem.* **2012**, *426*, 86–90. https://doi.org/10.1016/j.ab.2012.04.003.

35. Stone, T.A.; Cole, G.B.; Ravamehr-Lake, D.; Nguyen, H.Q.; Khan, F.; Sharpe, S.; Deber, C.M. Positive Charge Patterning and Hydrophobicity of Membrane-Active Antimicrobial Peptides as Determinants of Activity, Toxicity, and Pharmacokinetic Stability. *J. Med. Chem.* **2019**, *62*, 6276–6286. https://doi.org/10.1021/acs.jmedchem.9b00657.

36. Piovesana, S.; Capriotti, A.L.; Cavaliere, C.; La Barbera, G.; Montone, C.M.; Zenezini Chiozzi, R.; Laganà, A. Recent Trends and Analytical Challenges in Plant Bioactive Peptide Separation, Identification and Validation. *Anal. Bioanal. Chem.* **2018**, *410*, 3425–3444. https://doi.org/10.1007/s00216-018-0852-x.

37. Giacometti, J.; Buretić-Tomljanović, A. Peptidomics as a Tool for Characterizing Bioactive Milk Peptides. *Food Chem.* **2017**, *230*, 91–98. https://doi.org/10.1016/j.foodchem.2017.03.016.

Chapter IV

A complete workflow for discovering small bioactive peptides in foods by LC-MS/MS: A case study on almonds

This chapter was published as a journal article "**Huang, Y.-P.**; Dias, F.F.G.; de Moura Bell, J.M.L.N.; Barile, D. *Food Chem.* 2022, 369, 130834, doi:10.1016/j.foodchem.2021.130834."

Abstract

Identification of bioactive peptides is an increasingly important target for food chemists, particularly in consideration of the widespread application of proteolytic enzymes in food processing. Because the characterization of small peptides by LC-MS/MS is challenging, we optimized a dimethyl labeling technique to facilitate small peptide identification, using almond proteins as a model. The method was validated by comparing the MS/MS spectra of standards and almond-derived peptides in their nonderivatized and derivatized forms. Signal enhancement of a_1 ions was proved to effectively aid in the full-length sequencing of small peptides. We further validated this method using two industrially-relevant protein-rich extracts from almond flour: 1737 medium-sized peptides (5–39 amino acids) and 843 small peptides (2–4 amino acids) were identified. The use of an online bioactive peptide database, complemented by the existing literature, allowed the discovery of 208 small bioactive peptides, whereas for medium-sized peptides, only one was reported being bioactive.

Keywords: peptidomics, dimethyl labeling, di- and tripeptides, almonds (*Prunus dulcis*), bioactive peptides, enzymatic protein hydrolysis

4.1. Introduction

Food proteins represent one macronutrient that serves as a source of calories and provides the required building blocks for protein synthesis and for maintaining healthy human body functions. Aside from the nutritional value, some intermediate products formed during food protein hydrolysis may possess unique activities beneficial to human health, such as antihypertensive, antioxidant, and immunomodulatory activities, and are termed bioactive peptides (Korhonen & Pihlanto, 2006; Maestri et al., 2016). Proteolysis is known to naturally occur during gastrointestinal digestion, but can also be carried out in industrial settings by using fermentation and enzymatic treatments (Rios-Villa et al., 2020; Sanjukta & Rai, 2016; Vásquez-Villanueva et al., 2019).

Almonds (*Prunus dulcis*) are widely cultivated seeds harvested from the identically named tree, with a massive production worldwide (3.18 million tonnes in 2018), and have long been consumed in many countries for their nutritional benefits (*FAOSTAT Database*, 2020). As the demand for plant proteins has been rapidly increasing in recent years, due to growing health and environmental concerns linked to the consumption of animal products (meat and dairy), the market for almond products has been booming. Although tree nuts are considered major food allergens, epidemiological and clinical studies revealed that over 99% of consumers in various age groups are not allergic to almonds (Gupta et al., 2019; McWilliam et al., 2019; Sasaki et al., 2018). Therefore, considering almonds' massive production and uniquely high protein content (19.35-21.15%) (*FoodData Central*, 2020), almonds represent an ideal source of alternative proteins. However, unlike other protein-rich foods, such as dairy, seafood, and legumes (Korhonen & Pihlanto, 2006; Lafarga & Hayes, 2017; Maestri et al., 2016; Sanjukta & Rai, 2016), almonds' bioactive peptides have only been minimally studied and thus deserve more investigations.

Enzymatic hydrolysis using proteolytic enzymes has increasingly found applications in food processing for various purposes, including increasing protein extraction rate, enhancing protein functionality, increasing protein digestibility, and reducing allergenicity (Bahna, 2008; de Souza et al., 2020; Eriksen, 1983; Koopman et al., 2009). It also has been proven to be an effective strategy to generate bioactive peptides safely and at scale. Bioactive peptide generation via enzymatic hydrolysis was demonstrated on various foods and, in recent years, on some *Prunus* genus plant seeds (Lacroix & Li-Chan, 2012; Lafarga & Hayes, 2017; Liu et al., 2016; Vásquez-Villanueva et al., 2019).

The desirable bioactive properties of peptides are strictly determined by their structures. To discover bioactive peptides from foods and identify their structures, conventionally, a bioactivity-guided screening is performed after several fractionation steps based on peptide size, charge, and hydrophobicity. The structures of bioactive peptides in the fractions with the highest bioactivity can then be determined by Edman degradation (Matsui et al., 1999; Nakamura et al., 1995). Edman degradation has long been used for peptide sequencing following the bioactivity-guided screening (Matsui et al., 1999; Nakamura et al., 1995), and importantly, two of the most extensively studied bioactive peptides in dairy products, Val-Pro-Pro and Ile-Pro-Pro, were first identified in sour milk through this approach (Nakamura et al., 1995). However, due to the several limitations of using Edman degradation—including the requirement of individual peptide isolation and the lower throughput of peptide sequencing—peptide structure determination is nowadays more frequently accomplished by LC-MS/MS (Vásquez-Villanueva et al., 2019; Wang et al., 2017).

LC-MS/MS followed by automated data analysis using protein sequence database search can achieve high-throughput peptide characterization. A substantial number of bioactive peptide

sequences have been discovered in various types of foods using the bioactivity-guided screening, and they have been compiled in a few consistently curated freely available online databases, such as BIOPEP-UWM (Minkiewicz et al., 2019). With the bioactive peptide databases and comprehensive peptidomic profiles, theoretically, the identification of bioactive peptides could be easily achieved for novel food samples—if the peptide sequences matched the ones in the databases. However, another hurdle in this process is that peptide identification using the conventional workflow of LC-MS/MS analysis—followed by database search—mainly focuses on medium-sized peptides containing 5–20 amino acid residues. Medium-sized peptides, generated by digestion using specific proteolytic enzymes, are ideal for proteomics and are used for the identification and relative quantification of proteins. Nonetheless, when the goal is to study bioactive peptides, the conventional LC-MS/MS workflow will fail to identify a considerable number of bioactive peptide sequences with short lengths. Previous studies suggested that the potency of bioactive peptides and the enteral absorption efficiency are inversely related to peptide lengths (Hong et al., 2016; Roberts et al., 1999). Several food-derived small peptides were reported exhibiting specific bioactivities in animal models and human subjects (Dias et al., 2017; Li et al., 2019; Suetsuna et al., 2004). Some of those peptides, such as the dairy peptides Val-Pro-Pro and Ile-Pro-Pro, have already found significant applications in functional food products (Korhonen & Pihlanto, 2006; Lafarga & Hayes, 2017). Given the above, to explore the presence of bioactive peptides in particular foods and agricultural streams in a more comprehensive manner, it is necessary to study not only medium-sized peptides but also small peptides. Hence, there is a need to optimize and improve the conventional LC-MS/MS peptide characterization workflow.

Identification of amino acid sequences in small peptides can be achieved by *de novo* sequencing from LC-MS/MS data, in which the peptide sequence is elucidated from the sequential

assignment of fragment ions from a mass spectrum. However, analysis of small peptides using *de novo* sequencing for complex samples, such as a food matrix, can lead to ambiguous interpretations as each tandem MS spectrum often corresponds to two or more possible peptide sequences. A feasible solution we identified comes from the field of proteomics in which dimethyl labeling had been used for relative quantification (Hsu et al., 2003). An added advantage of this approach is that it generates more complete fragmentation patterns in the tandem MS spectra (Hsu et al., 2003, 2005) which in turn can aid in reducing the ambiguous interpretations in *de novo* sequencing of small peptides.

The present study investigated the potential of utilizing the dimethyl labeling technique to achieve small peptide identification and characterized the small and medium-sized peptides present in two industrially relevant protein-rich extracts from almond flour. Specifically, LC-MS/MS was used to analyze dimethyl-labeled samples in parallel with the corresponding nonderivatized samples, to accomplish comprehensive small peptide characterization. The obtained peptide sequences were then used for discovering potential activities through bioactive peptide database matching.

4.2. Materials and methods

4.2.1. Materials

Two protein-rich extracts were generated from almond flour (kindly donated by Blue Diamond Growers, Sacramento, CA) via aqueous extraction with and without using a protease at a pilot scale as described previously (de Almeida et al., 2019; Dias et al., 2020) and are referred to as ENZ and CTRL, respectively, in this work. Briefly, the extraction was conducted in a 10 L jacketed glass reactor model CG-1965-610M (Chemglass Life Sciences LLC, Vineland, NJ) with a 1:10 solids-to-liquid ratio at 50°C and pH 9 and stirring at 120 rpm for 60 min. For the aqueous

extraction with a protease, 0.5% (w/w) Neutral Protease 2 million from *Bacillus subtilis* (optimal temperature: 30–70°C, optimal pH: 5.5–9.0; BIO-CAT, Virginia, NY), which randomly cleaves peptide bonds in protein structures, was added to assist the extraction. The slurry was further separated into three fractions—the insoluble fraction, protein-rich extract, and oil-rich fraction. Peptide standards (V5626 and H2016), trifluoroacetic acid (TFA), formaldehyde solution, and ammonia solution 25% (LC-MS grade) were purchased from Sigma (St. Louis, MO). Acetonitrile (LC-MS grade) and formic acid (LC-MS grade) were purchased from Fisher Scientific (Waltham, MA). Sodium cyanoborohydride was purchased from TCI (Tokyo, Japan).

4.2.2. Peptide purification

4.2.2.1. Protein precipitation

The protein-rich extracts from almond flour (200 μL) were mixed with 400 μL ethanol and incubated at -30°C for 1 h for protein precipitation. After centrifugation (13,000g, −4°C, 30 min), the supernatants were transferred to new tubes and dried in a centrifugal evaporator (MiVac Quattro, Genevac Ltd., Ipswitch, Suffolk, UK) at 30 °C.

4.2.2.2. Solid-phase extraction

After removing proteins by precipitation, the dried samples were re-dissolved with nanopure water and further purified with solid-phase extraction (SPE). The re-dissolved samples were premixed with equal volumes of 0.2% TFA (v/v) and loaded to SPE cartridges (Discovery DSC-18, 500 mg, 3 mL tube, Sigma) preconditioned with 5 mL of acetonitrile and 5 mL of 0.1% TFA (v/v). The SPE cartridges were then washed with 6 mL of 0.1% TFA for eliminating salts and polar interferences and eluted with 6 mL of 80% acetonitrile containing 0.1% TFA (v/v/v) for collecting peptides. The eluates containing peptides were dried in a centrifugal evaporator at 30°C, re-dissolved and appropriately diluted with 2% acetonitrile, and analyzed with LC-MS/MS.

4.2.3. Dimethyl labeling via reductive amination

Dimethyl labeling was performed according to Hsu et al. (2003) with slight modification on peptide standards and the protein-rich extracts from almond flour that underwent protein precipitation. Briefly, peptide standards (50 μg/mL) and the almond protein-rich extract samples (equivalent to 100 μL of the original protein-rich extract) dissolved in 100 μL sodium acetate buffer (100 mM, pH 5.5) were mixed with 10 μL 4% formaldehyde and 10 μL 1 N sodium cyanoborohydride and incubated at room temperature for 1 h. After the labeling reaction, 10 μL 5% ammonia solution was added to consume the excess formaldehyde. The samples were then acidified by adding 4 μL TFA and purified by C18 SPE as described in section 4.2.2.2. Purified dimethyl-labeled samples were analyzed with LC-MS/MS in parallel to the corresponding nonderivatized samples.

4.2.4. 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). The nonderivatized and dimethyl-labeled samples were delivered to the enrichment column (360 nL) of an Agilent Polaris-HR-high-capacity Chip 3C18 with 3% acetonitrile, 0.1% TFA in water (v/v/v) at a flow rate of 2.5 μL min^{-1} and further flushed with 1 μL of the eluent. Chromatographic separation of peptides was performed with a gradient consisted of eluents of 3% acetonitrile, 0.1% TFA in water (v/v/v; A) and 90% acetonitrile, 0.1% TFA in water (v/v/v; B) at a flow rate of 0.3 μL min^{-1}. The gradient was linearly increased from 0% B to 30% B in 40 min and further increased to 45% B in 5 min, followed by flushing with 100% B for 5 min and equilibrating at 100% A for 15 min before the next injection. The capillary voltage was set at 1880 V. The drying gas was set at 325°C with a flow rate of 5 L min^{-1}. Mass-charge-ratio (*m/z*) was scanned at a rate of 8 spectra sec^{-1} in the range

of 130-1400 for nonderivatized samples and 150-1400 for dimethyl-labeled samples. The precursors were then selected based on abundances and isolated with a width of 1.3 amu for fragmentation. A ramped collision energy with the equation ($0.03 \times m/z + 2$) was applied to ions of any charges for nonderivatized samples. For dimethyl-labeled samples, collision energy with the equations ($0.04 \times m/z + 2$) and ($0.03 \times m/z + 5$) was used for singly and multiply charged ions, respectively. The MS/MS analysis was scanned at a rate of 0.8 spectra sec^{-1}.

4.2.5. Data analysis for peptide identification

PEAKS Studio X+ (Bioinformatics Solutions Inc., Waterloo, ON, Canada) was used for analyzing LC-MS/MS data for peptide identification. Peptides with 2–4 amino acid residues were identified through automatic *de novo* sequencing from both the nonderivatized and dimethyl-labeled samples. All the *de novo* matches of di-, tri-, and tetrapeptides with Denovo scores ≥ 20 in PEAKS Studio X+ were manually inspected to find the correct identification from the candidate matches, and only peptides that were confidently identified with full-length sequencing were taken. Peptides with ≥ 5 amino acid residues were identified from nonderivatized samples through database search with proteins in the Uniprot database (https://www.uniprot.org/, accessed 3/10/2020), both Swiss-Prot and TreMBL, with the organism name *Prunus dulcis*. Only the peptide sequences listed in the "top" proteins in the PEAKS results were included in the final medium-sized peptide identification list. For both *de novo* sequencing and database search, the variable modifications included deamidation (N and Q), phosphorylation (S, T, and Y), and oxidation (M) for both nonderivatized and dimethyl-labeled samples. The following modifications were exclusively applied to dimethyl-labeled samples: dimethyl labeling (28.0313 Da) on Lys at any position and all N-terminal amino acids except for Pro and Lys, monomethyl labeling (14.0156 Da) on N-terminal Pro, and double dimethyl labeling (56.0626 Da) on N-terminal Lys. All the

modifications resulting from dimethyl labeling were set as fixed in PEAKS with the sole exception of “dimethyl labeling on Lys at any position”, which was set as variable to exclude the circumstance of double dimethyl labeling on N-terminal Lys. During the subsequent manual validation, peptide sequences containing Lys were only considered if the Lys residues were modified with either double dimethyl labeling at the N-terminus or dimethyl labeling at other positions. Mass error tolerance was set at 15 ppm and 0.02 Da for precursors and fragments, respectively.

4.2.6. Database matching for bioactive peptide identification

For bioactive peptide identification, the list of peptides sequences identified from the LC-MS/MS results were searched against bioactive peptide sequences in the BIOPEP-UWM database (http://www.uwm.edu.pl/biochemia/index.php/en/biopep, accessed 8/10/2020). As small peptides were identified *de novo*, and Leu and Ile are isomeric and indistinguishable in the LC-MS/MS analysis, all possible sequences of small peptide identifications containing Leu or Ile were included in the bioactive peptide database matching. The BIOPEP-UWM database was selected because it compiles bioactive peptide data from various food sources and synthetic peptides, with several different activities, and it is routinely updated with new information from the literature.

4.3. Results and discussion

4.3.1. Characterization of small peptide (2–4 amino acid residues)

4.3.1.1. Using dimethyl labeling to facilitate small peptide identification: technical advantages and limitations

Typically, automatic data analysis with database search can identify peptides consisting of five amino acids or larger. Small peptides identification theoretically can be accomplished by *de novo* sequencing, which currently can also be done automatically with specific software for more

efficient analysis (Zhang et al., 2012). However, *de novo* sequencing of small peptides may result in several possible peptide sequences from one single spectrum, leaving the researcher with multiple possible interpretations (Murray et al., 2018). For example, isomers comprising the same amino acids in different orders are frequently seen. To manage this issue, in this study, we attempted to use dimethyl labeling, which is known for signal enhancement of a_1 ions in tandem MS analysis (Hsu et al., 2005), and hence can improve the identification of N-terminal residues.

We first tested a couple of peptide standards, including Val-Tyr and Val-Tyr-Val, to gain knowledge on the fragmentation pattern of dimethyl-labeled small peptides. The results showed that the fragmentation pattern changed substantially after dimethyl labeling. For the nonderivatized dipeptide Val-Tyr, Val and Tyr immonium ions and y_1 ion were observed in tandem MS analysis; for the dimethyl-labeled Val-Tyr, the predominant a_1 ion was the only fragment observed (Fig. 4.1A). The fragments observed for nonderivatized Val-Tyr-Val included the Val and Tyr immonium ions, a_2, b_2, and y_2 ions. After dimethyl labeling, the a_1 ion peak dominated the tandem MS spectrum, and the a_2 and b_2 ions were also detected but in significantly lower abundances than the a_1 ion (Fig. 4.1B). Both Val-Tyr and Val-Tyr-Val formed singly charged ions after ionization. Dimethyl labeling increased the basicity of the N-terminal Val on the two peptide standards and made them ready to accept the only proton. Therefore, the proton favored remaining on the N-terminal fragments to generate the enhanced a_1 ions and other a and b ions during fragmentation, whereas the C-terminal fragments turned out to be neutral and thus undetectable (Locke et al., 2006).

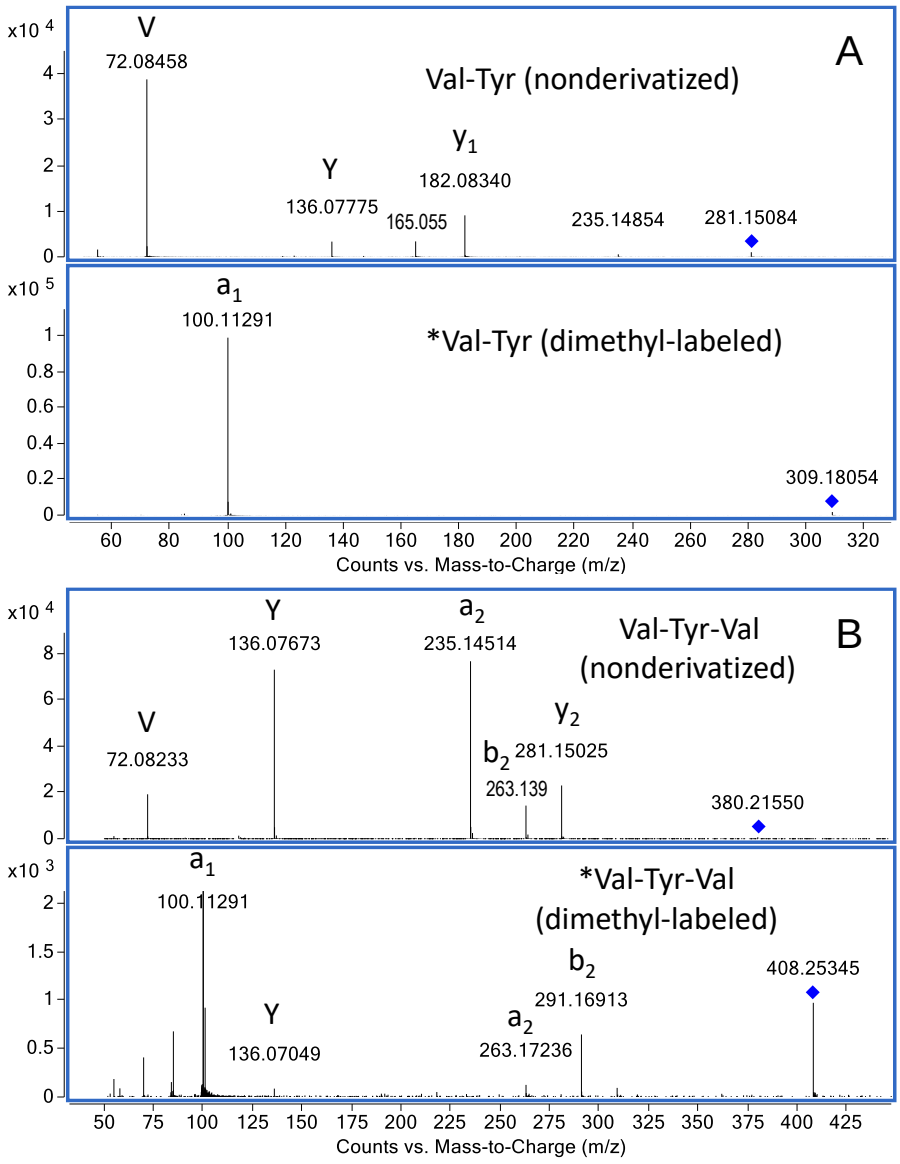


Fig. 4.1. Q-TOF tandem MS spectra of nonderivatized and dimethyl-labeled peptide standards (A. Val-Tyr; B. Val-Tyr-Val). The star symbol denotes the positions being dimethyl-labeled. The spectrum of dimethyl-labeled Val-Tyr-Val was zoomed in; the peak intensity of the a_1 ion of dimethyl-labeled Val-Tyr-Val was 9.0×10^5.

After testing peptide standards, we applied dimethyl labeling to the almond ENZ protein-rich extract, which was extracted from almond flour in the presence of neutral protease and was

expected to contain a significant number of small peptides. Hence, additional peptide sequences were available for further studying the fragmentation pattern of dimethyl-labeled small peptides and analyzing their differences from the nonderivatized peptides, thus gaining more information about our method's merits. In general, the majority of dimethyl-labeled small peptides formed singly charged ions and had similar fragmentation patterns with the two peptide standards described above. Their tandem MS spectra mainly contained a and b ions, and depending on the peptide sequences, y ions were sometimes observed. The apparent spectral peaks with the lowest *m/z* on the tandem MS spectra were typically the a_1 ions. In comparison, singly charged nonderivatized peptides usually did not exclusively generate a and b ions nor form apparent a_1 ions. Peptide sequences containing basic amino acids, such as Arg and Lys, usually formed doubly charged ions. Because typically, both the N-terminal and C-terminal fragments could get a proton after fragmentation, both a and b ions and y ions were seen in their tandem MS spectra.

Overall, dimethyl labeling resulted in a great deal of benefits for small peptide identification. The motivation for using dimethyl labeling in this study was to utilize apparent a_1 ions to aid the identification of the amino acid residues at N-termini and to fulfill full-length sequencing of small peptides. This approach was indeed proved to be effective in the identification of several small peptides. For example, in Fig. 4.2A, a tetrapeptide in the ENZ protein-rich extract might correspond to Gly-Val-Lxx-Tyr or Val-Gly-Lxx-Tyr (Lxx represents Leu or Ile, which are isomeric and were indistinguishable in the LC-MS/MS analysis) when analyzed nonderivatized because the y_3 ion was not detected in the tandem MS analysis. In contrast, when it was dimethyl-labeled, an apparent a_1 ion appeared on the tandem MS spectrum; thus, the tetrapeptide sequence could be confidently determined as Gly-Val-Lxx-Tyr. The a_1 ion formed by nonderivatized Gly typically cannot be detected due to the low mass (*m/z* 30.03). In comparison, the dimethyl-labeled

N-terminal Gly formed an apparent a_1 ion with a unique mass (*m/z* 58.07), which would not be confused with immonium ions originated from other locations in the peptide sequence, and thus enabled the identification of the N-terminal amino acid residue as well as the full-length sequencing.

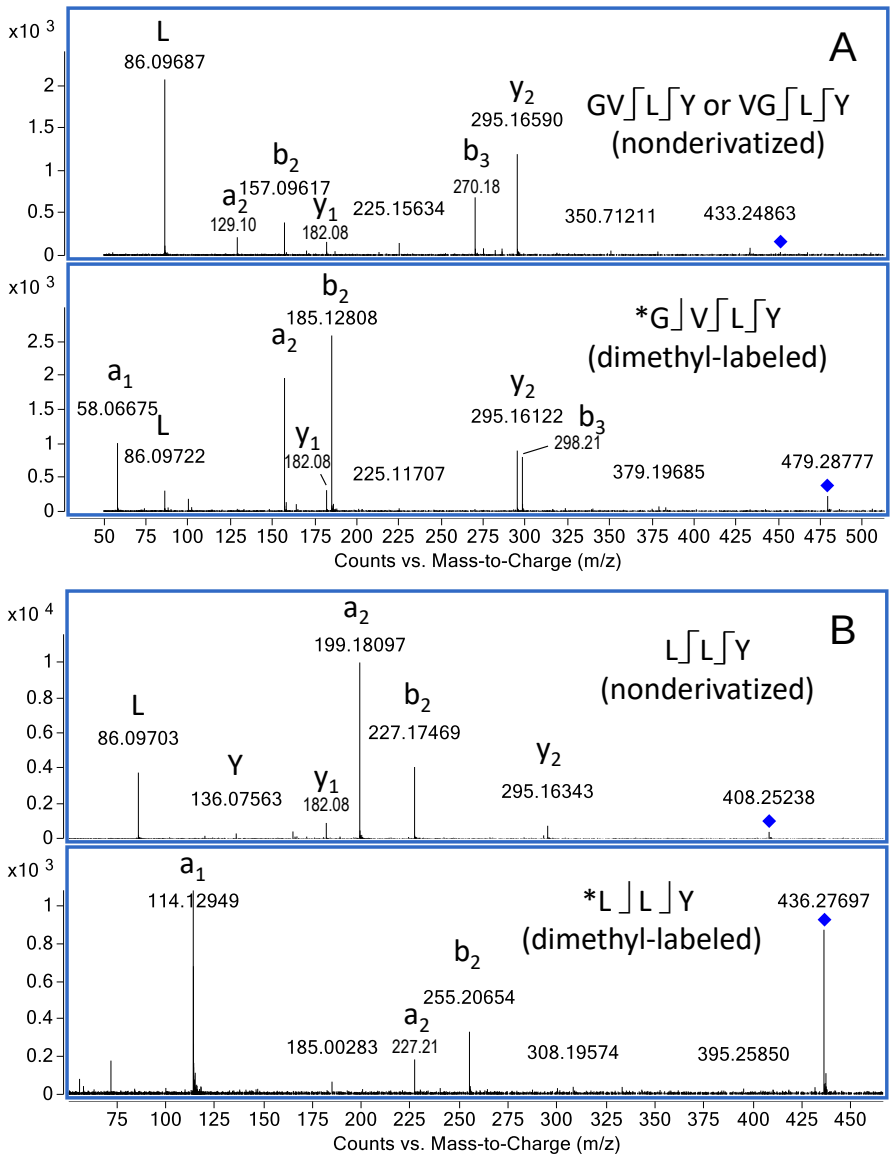


Fig. 4.2. Q-TOF tandem MS spectra of nonderivatized and dimethyl-labeled peptide identified from the ENZ protein-rich extract from almond flour (A. Gly-Val-Lys-Tyr; B. Lxx-Lxx-Tyr). The star symbol

denotes the positions being dimethyl-labeled. The spectrum of dimethyl-labeled Lxx-Lxx-Tyr was zoomed in; the peak intensity of the a_1 ion of dimethyl-labeled Lxx-Lxx-Tyr was 1.4×10^4.

Moreover, to ensure *de novo* sequencing's correctness, manual validation was performed after the automatic data analysis using PEAKS. In most cases, dimethyl labeling simplified the tandem MS spectra of singly charged peptides by filtering out most y ions and thus facilitated manual validation. Furthermore, dimethyl labeling increased the hydrophobicity of (originally) hydrophilic peptides as well as their signal intensities. This consequently improved the data quality of tandem MS spectra and enabled identifying a higher number of hydrophilic peptides.

Some limitations of small peptide identification with dimethyl-labeled peptides were also recognized. In the analysis of small peptides, dimethyl labeling often led to lower Denovo scores in PEAKS (e.g., Denovo scores of dimethyl-labeled peptides were on average 66% (33–178%) of the corresponding nonderivitized peptides for small peptides in the almond ENZ protein-rich extract; Supplementary material Table 4.S1). As mentioned above, tandem MS spectra of singly charged peptide ions contained mainly a and b ions. The lack of y ions reduced the number of matched fragments and therefore decreased the peptide-spectrum matching scores. Although this problem can be solved using manual validation for rescuing correct identifications with lower matching scores, further study in modifying sequencing algorithms will be needed to improve the correctness in automatic data analysis and increase the analysis throughput.

Besides, in tandem MS analysis of singly charged dimethyl-labeled peptides, signal weakening of fragments other than a_1 ions was frequently seen and particularly evident for peptides with bulky and hydrophobic amino acid, such as Leu, Ile, and Phe, on N-termini. For example,

Lxx-Lxx-Tyr, a tripeptide identified in the ENZ protein-rich extract, generated more abundant a_2 and b_2 ions than the a_1 ion when it was nonderivatized, but after dimethyl labeling, the abundances of a_2 and b_2 ions were both less than 5% of the a_1 ion (Fig. 4.2B). In some cases, the signal weakening of a and b ions other than a_1 ions even restricted the sequencing of amino acid residues near C-termini. Fortunately, it is still possible to identify these peptides with the nonderivatized form, as long as the y_{n-1} (n = number of amino acid residues in peptides) ions were observed on tandem MS spectra. Therefore, analyzing nonderivatized and dimethyl-labeled peptides in parallel supposedly would maximize the number of small peptide identification.

Although the majority of previous peptidomic studies only focused on the analysis of medium-sized peptides (Agyei et al., 2018; Jin et al., 2016; Rios-Villa et al., 2020), peptidomic profiling, including both small and medium-sized peptides, is crucial for more precisely delineating the properties of various types of foods, such as enzymatic hydrolysates and fermented foods, and even more crucial to understand food digestion using simulated gastrointestinal systems, which are now commonly employed in research labs (Egger et al., 2016; Rios-Villa et al., 2020). Enzymatic hydrolysis, fermentation, and gastrointestinal digestion are all potential routes for generating bioactive peptides. As proteolysis gets more extensive, greater amounts of small peptides are expected to be generated, and thus small peptides should be included in peptidomic analysis instead of being neglected. Dimethyl labeling is an easy-to-use, rapid, and affordable approach to facilitate full-length sequencing of certain types of small peptides. The method introduced in the present work - analyzing nonderivatized and dimethyl-labeled small peptides in parallel - provided valuable information to complement the conventional peptidomic profiling for medium-sized peptides, achieved by LC-MS/MS and the subsequent data analysis with database search, to allow more comprehensive peptidomic profiling.

4.3.1.2. Small peptide identification in protein-rich extracts from almond flour

4.3.1.2.1. Small peptides in protein-rich extracts

In a previous study, our group applied enzyme-assisted extraction, which is an environmentally friendly strategy, to simultaneously extract proteins and lipids from almond flour (de Almeida et al., 2019). It is expected that a considerable amount of peptides, potentially along with some bioactive peptides, were generated during the extraction. By analyzing nonderivatized and dimethyl-labeled peptides in parallel, 753 and 219 small peptides comprising 2–4 amino acid residues were identified in the ENZ and CTRL protein-rich extracts, respectively (Fig. 4.3; Supplementary data Table 4.S1 and 4.S2). The peptides present in the CTRL are expected to be either naturally occurring (already found in the almond flour) or formed during the process. Therefore, the peptides truly generated by proteolysis can be identified by comparing the list of peptides found in the ENZ and CTRL samples. Not surprisingly, the number of small peptides identified from the ENZ was considerably higher than the CTRL one, indicating that small peptides encrypted in proteins or longer peptides were efficiently released under the action of neutral protease. Among the identified small peptides, 129 peptide sequences were identified in both the ENZ and CTRL (overlapping sequences in Supplementary material Table 4.S1 and 4.S2). For this comparison, due to minor shifts in retention times between the ENZ and CTRL samples, when peptide sequence identities could not be unequivocally determined because of the presence of isomeric Leu/Ile, we assumed that the peptides were identical. This implied that some small free peptides in the starting material were further broken down by the protease during extraction, therefore, solely identified from the CTRL protein-rich extract, whereas some small peptides might have survived proteolysis and were found in both the ENZ and CTRL protein-rich extracts.

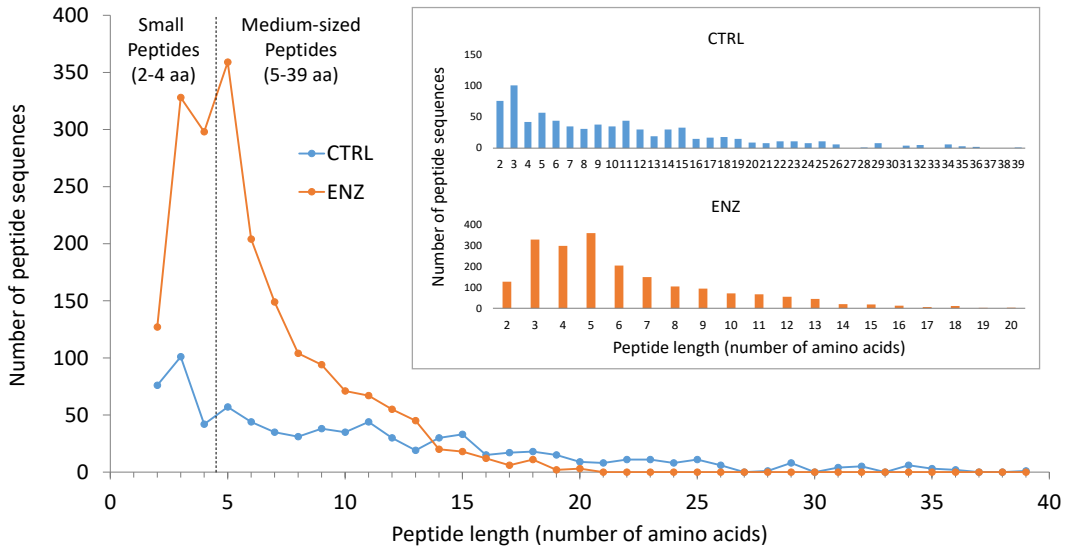


Fig. 4.3. Length distribution of peptides identified in the protein-rich extracts from almond flour. Peptides composed of 2–4 amino acids were identified via *de novo* sequencing from both nonderivatized and dimethyl-labeled samples; peptides with ≥5 amino acids were identified through database search from a nonderivatized sample.

4.3.1.2.2. Advantages of dimethyl labeling for small peptide analysis: increased number of identification in protein-rich extracts from almond flour

The Venn diagram in Fig. 4.4 graphically illustrates the comparison of nonderivatized and dimethyl-labeled samples in terms of small peptide sequences identified in the protein-rich extracts from almond flour. For the ENZ protein-rich extract, 420 and 540 peptide sequences were identified from the nonderivatized and dimethyl-labeled samples, respectively, with 207 common sequences; for the CTRL protein-rich extracts, 121 and 158 peptide sequences were identified in the nonderivatized and dimethyl-labeled samples, respectively, and 61 sequences were identified

from both. The relatively small overlapping between the peptide sequences identified from the nonderivatized and dimethyl-labeled samples revealed that using both of the two approaches effectively increased the number of small peptide identification.

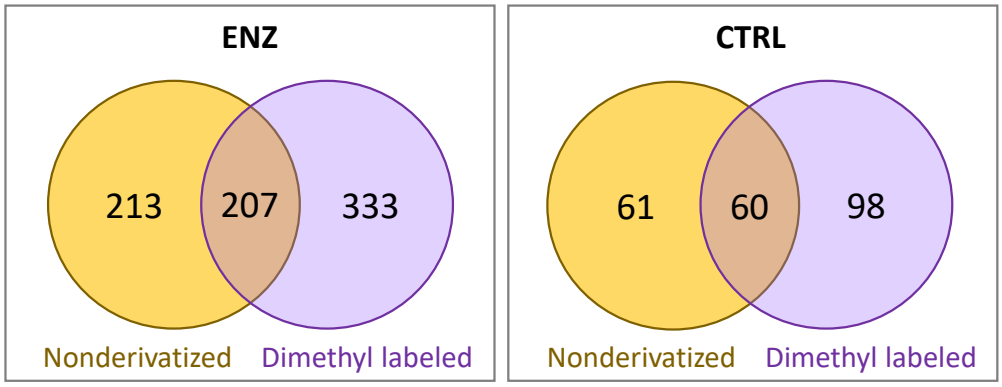


Fig. 4.4. Venn diagrams of numbers of small peptide sequences identified in the nonderivatized and dimethyl-labeled samples of the protein-rich extracts from almond flour.

Notably, the approach of dimethyl labeling was particularly beneficial for identifying peptides with an N-terminal Ala or Gly. For example, for the ENZ protein-rich extract, 15 and 5 small peptide sequences with Ala and Gly on the N-termini, respectively, were identified in the nonderivatized sample, whereas 44 and 54 small peptide sequences with Ala and Gly, respectively, on the N-termini were identified in the dimethyl-labeled sample (Supplementary data Table 4.S1). This further confirmed that the apparent a_1 ions of dimethyl-labeled Ala and Gly efficiently aided in small peptides' full-length sequencing.

4.3.2. Characterization of medium-sized peptides (≥ 5 amino acid residues) in protein-rich extracts from almond flour

4.3.2.1. Medium-sized peptide identification

Peptidomic analysis for medium-sized peptides in the protein-rich extracts from almond flour was performed by LC-MS/MS followed by data analysis using database search. Through this approach, 1219 and 554 medium-sized peptides were identified in the ENZ and CTRL, respectively (Supplementary data Table 4.S3 and 4.S4). The ENZ protein-rich extract had more than double number of peptide sequences compared to the CTRL one. Among the over 1700 medium-sized peptides identified in the ENZ and CTRL, only 36 medium-sized peptide sequences were found in both the protein-rich extracts, indicating that medium-sized peptides identified in the ENZ were mostly generated by the neutral protease. On the other hand, the majority of the medium-sized peptides identified in the CTRL were not found in the ENZ, indicating that most of those peptides had been broken down into smaller peptides (or even free amino acids) by the neutral protease. Noteworthy, the length distribution of the identified peptides was rather distinct between the ENZ and CTRL (Fig. 4.3). The CTRL contained peptides with chain lengths as large as 39 amino acid residues, whereas the longest peptide sequences identified from the ENZ were only 20 amino acid long. Also, peptide sequences with shorter lengths primarily dominated the ENZ compared to the CTRL. Peptide sequences composed of 5 amino acid residues accounted for 29% and 10% of the medium-sized peptide sequences identified in the ENZ and CTRL, respectively, and peptide sequences composed of 5-10 amino acids represented 80% and 43% of the medium-sized peptide sequences, respectively.

4.3.2.2. Major almond proteins generating medium-sized peptides

The most important parent proteins generating the medium-sized peptides were identified as prunin-1 and prunin-2 in both the protein-rich extracts, no matter whether the protease was used or not (Supplementary Material Table 4.S5). This was in line with the fact that amandin, a hexameric protein composed of prunin-1 and prunin-2, accounts for near 70% of the total protein of almonds (Costa et al., 2012; Wolf & Sathe, 1998). The number of peptide sequences that originated from several proteins, such as prunin-2, vicilin, and (R)-mandelonitrile lyase, was much higher in the ENZ protein-rich extract than in the CTRL one (Supplementary data Table 4.S5). Peptide sequences originating from certain proteins, such as NAD(P)-binding Rossmann-fold superfamily protein, were exclusively identified in the ENZ protein-rich extract.

4.3.3. Bioactive peptides identification from almond protein-rich extracts

4.3.3.1. Identification of bioactivity for small peptides

From the ENZ and CTRL protein-rich extracts, 188 and 110, respectively, bioactive peptide sequences comprising 2–4 amino acid residues were found (Table 4.1). Peptide sequences of 2–4 amino acids are often ubiquitous. Identical sequences could frequently found in different proteins and even in proteins from unrelated organisms. Although studies about bioactive peptides derived from almonds are still scarce, a significant number of bioactive peptide sequence matches were still obtained for small peptides present in the two protein-rich extracted from almond flour. The majority of the identified small bioactive peptides had dipeptidyl peptidase (DPP) IV inhibitory (ENZ: 112; CTRL: 63), ACE inhibitory (ENZ: 106; CTRL: 61), antioxidative (ENZ: 23; CTRL: 14), DPP III inhibitory (ENZ: 23; CTRL: 12), and renin inhibitory (ENZ: 6; CTRL: 6) activities. Among those identified small bioactive peptides, several of them possessed 2–5 different activities (Table 4.1).

Table 4.1. Bioactivity identification for small peptides found in the protein-rich extracts from almond flour.

Peptide	Activities (BIOPEP ID; IC50)	ENZ	CTRL
AF	ACE inhibitor (7583; 190 μM), DPP IV inhibitor (8759)	✓	✓
AI	ACE inhibitor (8193; 3.41 μM)	✓	✓
AIP	ACE inhibitor (3597; 670 μM)		✓
AL	DPP IV inhibitor (8559; 882.13 μM)	✓	✓
ALP	ACE inhibitor (9029; 239.9 μM)		✓
ASL	ACE inhibitor (8968; 102.15 μM)	✓	
AV	ACE inhibitor (8951; 956.28 μM), DPP IV inhibitor (8764)	✓	✓
AVL	ACE inhibitor (9060; 7.11 μM)	✓	✓
AVP	ACE inhibitor (3370; 340 μM)	✓	✓
AW	ACE inhibitor (7543; 10 μM), antioxidative (8460), DPP IV inhibitor (8695)	✓	✓
AY	ACE inhibitor (3563; 19 μM), antioxidative (7866), DPP IV inhibitor (8765)	✓	
EI	ACE inhibitor (7826), DPP IV inhibitor (8772)	✓	✓
EL	antioxidative (7888)	✓	✓
EY	ACE inhibitor (7752; 2.68 μM), DPP IV inhibitor (8777)	✓	
FA	DPP IV inhibitor (3176), DPP III inhibitor (9500)	✓	
FAL	ACE inhibitor (7823; 26.3 μM)	✓	✓
FG	ACE inhibitor (7605; 3700 μM)	✓	
FL	DPP IV inhibitor (8555; 399.58 μM), DPP III inhibitor (9502)	✓	✓
FM	DPP III inhibitor (9503)	✓	
FN	DPP IV inhibitor (8778)	✓	
FQ	ACE inhibitor (9076; 51.29 μM), DPP IV inhibitor (8779)	✓	
FQP	ACE inhibitor (3341)		✓
FR	ACE inhibitor (7592; 920 μM), DPP IV inhibitor (8780), DPP III inhibitor (9501)	✓	✓
FT	renin inhibitor (2835)	✓	✓
FY	ACE inhibitor (3556; 25 μM)	✓	✓
GF	ACE inhibitor (7591; 630 μM), DPP IV inhibitor (8782), DPP III inhibitor (9488)	✓	✓
GFL	immunostimulating (3061), regulating (2737), DPP III inhibitor (9512)	✓	
GI	ACE inhibitor (7596; 1200 μM), DPP IV inhibitor (8785)	✓	✓
GL	ACE inhibitor (7599; 2500 μM), DPP IV inhibitor (8561; 2615.03 μM)	✓	✓
GLY	ACE inhibitor (9033; 8.91 μM), regulating (2739)	✓	
GM	ACE inhibitor (7597; 1400 μM)	✓	
GV	ACE inhibitor (7608; 4600 μM), DPP IV inhibitor (8786)	✓	✓
GVLY	ACE inhibitor (9325; 16 μM)	✓	
GVY	ACE inhibitor (9065; 398.1 μM)	✓	
GY	ACE inhibitor (3532; 210 μM), DPP IV inhibitor (8788)	✓	

HF	DPP IV inhibitor (8791), DPP III inhibitor (9495)	✓	
HI	DPP IV inhibitor (8793)	✓	
HL	ACE inhibitor (7602; 3200 μM), antioxidative (3317), DPP IV inhibitor (8557; 143.19 μM), DPP III inhibitor (9493)	✓	
HLL	ACE inhibitor (7479; 22.2 μM)	✓	
HV	DPP IV inhibitor (8797)		✓
HY	ACE inhibitor (3494; 26.1 μM), DPP IV inhibitor (8799)	✓	
IA	ACE inhibitor (7562; 153 μM), DPP IV inhibitor (8525)	✓	✓
IAK	ACE inhibitor (7626; 15.7 μM)	✓	✓
IAP	ACE inhibitor (7819; 2.7 μM)		✓
IAQ	ACE inhibitor (9043; 34.67 μM)	✓	✓
IAY	ACE inhibitor (9071; 12.59 μM)	✓	
IE	ACE inhibitor (7827)		✓
IF	ACE inhibitor (7593; 930 μM)	✓	✓
IFG	ACE inhibitor (7639; 1001 μM)		✓
IG	ACE inhibitor (7595; 1200 μM)	✓	✓
IH	DPP IV inhibitor (8800), DPP III inhibitor (9497)	✓	✓
II	stimulating (8325), DPP IV inhibitor (8801)	✓	✓
IIY	ACE inhibitor (9704; 1.08 μM)	✓	
IKK	antioxidative (7862)		✓
IKY	ACE inhibitor (8229; 34 μM)		✓
IL	ACE inhibitor (9079; 54.95 μM), stimulating (8323), DPP IV inhibitor (8802)	✓	✓
IM	DPP IV inhibitor (8803)	✓	✓
IN	DPP IV inhibitor (8804)	✓	✓
IP	ACE inhibitor (7581; 130 μM), DPP IV inhibitor (8501; 410 μM)		✓
IPI	DPP IV inhibitor (3167; 7.4 μM)	✓	
IPY	ACE inhibitor (7803; 206 μM)	✓	
IQ	DPP IV inhibitor (8805)	✓	✓
IQP	ACE inhibitor (8184; 3.8 μM), DPP IV inhibitor (8693)	✓	
IR	ACE inhibitor (3258; 695 μM), antioxidative (8215), DPP IV inhibitor (8806), CaMPDE inhibitor (8247), renin inhibitor (8246; 9200 μM)	✓	✓
ITF	ACE inhibitor (9062; 48.98 μM)	✓	
IV	stimulating (8322)	✓	✓
IVF	ACE inhibitor (8508; 63.3 μM)	✓	✓
IVG	HMG-CoA reductase inhibitor (9384)		✓
IVQ	ACE inhibitor (9045; 95.5 μM)	✓	✓
IVR	ACE inhibitor (7502; 0.81 μM)	✓	
IVY	ACE inhibitor (7541; 0.48 μM)	✓	
IW	ACE inhibitor (7544; 4.7 μM), DPP IV inhibitor (8807)	✓	
IY	ACE inhibitor (3383; 2.1 μM), antioxidative (7873)	✓	✓

IYK	ACE inhibitor (7656; 177 μM)	✓	✓
IYP	ACE inhibitor (2627; 61 μM)	✓	
IYPR	ACE inhibitor (3540; 10 μM)		✓
KF	ACE inhibitor (7692; 28.3 μM), DPP IV inhibitor (8809), CaMPDE inhibitor (8249)	✓	
KI	DPP IV inhibitor (8812)	✓	✓
KK	bacterial permease ligand (3751), DPP IV inhibitor (8813)	✓	
KL	ACE inhibitor (7693; 50.2 μM)	✓	✓
KP	ACE inhibitor (7810; 22 μM), antioxidative (8218), DPP IV inhibitor (8519; 2540 μM)		✓
KV	DPP IV inhibitor (8817)		✓
KY	ACE inhibitor (7691; 13 μM), DPP IV inhibitor (8819)	✓	
LA	ACE inhibitor (7585; 310 μM), activating ubiquitin-mediated proteolysis (4006), DPP IV inhibitor (3175), DPP III inhibitor (9499)	✓	✓
LAA	ACE inhibitor (3539; 13 μM)	✓	✓
LAP	ACE inhibitor (7570; 3.5 μM)		✓
LAY	ACE inhibitor (3558; 3.9 μM)	✓	
LF	ACE inhibitor (3551; 349 μM)	✓	✓
LG	ACE inhibitor (7619; 8800 μM)	✓	✓
LGI	ACE inhibitor (9061; 28.84 μM)	✓	✓
LH	antioxidative (3305), DPP IV inhibitor (8820)	✓	✓
LHF	antioxidative (7991)	✓	
LI	stimulating (8324), DPP IV inhibitor (8821)	✓	✓
LIY	ACE inhibitor (7657; 0.8 μM)	✓	
LK	antioxidative (8217)	✓	✓
LL	stimulating (8326), DPP IV inhibitor (3182)	✓	✓
LLF	ACE inhibitor (7807; 79.8 μM)	✓	✓
LLL	stimulating (3356)	✓	✓
LLPH	antioxidative (3314)	✓	
LLR	antioxidative (8484)	✓	✓
LLY	immunostimulating (3065)	✓	
LM	DPP IV inhibitor (8822)	✓	✓
LN	ACE inhibitor (7832), DPP IV inhibitor (8823)	✓	✓
LP	DPP IV inhibitor (3180; 2370 μM)		✓
LPF	ACE inhibitor (9040; 39.81 μM)	✓	
LPL	DPP IV inhibitor (8616; 241.4 μM)	✓	
LQ	ACE inhibitor (7831)	✓	✓
LQP	ACE inhibitor (3542; 1.9 μM), DPP IV inhibitor (8689; 1181.1 μM)	✓	
LQQ	ACE inhibitor (3714; 100 μM)		✓
LR	ACE inhibitor (9213; 158 μM), DPP III inhibitor (9478), renin inhibitor (2842)	✓	✓
LT	DPP IV inhibitor (8824)	✓	✓

LTF	ACE inhibitor (7638; 330 μM)	✓	
LV	stimulating (8321), DPP IV inhibitor (8825)	✓	✓
LVL	ACE inhibitor (3421; 12.3 μM)	✓	✓
LVQ	ACE inhibitor (9048; 14.13 μM)	✓	✓
LVR	ACE inhibitor (3528; 14 μM)	✓	
LVY	ACE inhibitor (8402; 5.84 μM)	✓	
LW	ACE inhibitor (3389; 50 μM), antioxidative (8462), DPP IV inhibitor (8688; 993.4 μM), DPP III inhibitor (9498), renin inhibitor (2832)	✓	
LY	ACE inhibitor (3381; 18 μM), antioxidative (7872), renin inhibitor (9470; 1870 μM)	✓	✓
LYP	ACE inhibitor (7550; 6.6 μM)	✓	
MF	ACE inhibitor (3385; 45 μM), DPP IV inhibitor (8827)	✓	
MI	DPP IV inhibitor (8830)	✓	✓
MK	DPP IV inhibitor (8831)	✓	✓
ML	DPP IV inhibitor (8832; 91 μM)	✓	✓
MM	ACE inhibitor (9085; 547.5 μM), antioxidative (9086; 547.5 μM), DPP IV inhibitor (8833; 93 μM)	✓	
NL	DPP IV inhibitor (8845)	✓	✓
NLR	ACE inhibitor (9754)	✓	
NM	DPP IV inhibitor (8846)	✓	
NY	ACE inhibitor (7682; 32.6 μM), DPP IV inhibitor (8853)	✓	
PF	DPP IV inhibitor (8854), DPP III inhibitor (9505)	✓	
PI	DPP IV inhibitor (8857)		✓
PL	ACE inhibitor (7513; 337.32 μM), DPP IV inhibitor (8638)		✓
PQ	ACE inhibitor (7837), DPP IV inhibitor (8861)	✓	✓
PR	ACE inhibitor (3537; 4.1 μM), DPP III inhibitor (9489)	✓	✓
PT	ACE inhibitor (7833), DPP IV inhibitor (8863)	✓	
PW	antioxidative (8190), DPP IV inhibitor (8865)	✓	
QI	DPP IV inhibitor (8873)	✓	✓
QL	DPP IV inhibitor (8874)	✓	✓
QQ	DPP IV inhibitor (8876)	✓	
QY	DPP IV inhibitor (8881)	✓	✓
RIY	ACE inhibitor (7821; 28 μM)	✓	
RVR	ACE inhibitor (9327; 526 μM)	✓	
SF	ACE inhibitor (7685; 130.2 μM), DPP IV inhibitor (8891), renin inhibitor (9432)		✓
SI	DPP IV inhibitor (8893)	✓	✓
SL	DPP IV inhibitor (8560; 2517.0801 μM)	✓	✓
SV	DPP IV inhibitor (8895)	✓	
SW	DPP IV inhibitor (8896)	✓	
SY	ACE inhibitor (7684; 66.3 μM), DPP IV inhibitor (8897)	✓	

TF	ACE inhibitor (8185; 18 μM), DPP IV inhibitor (8900), DPP III inhibitor (9486), renin inhibitor (9471; 3100 μM)	✓	✓
TI	DPP IV inhibitor (8903)	✓	✓
TK	DPP IV inhibitor (8904)	✓	✓
TL	DPP IV inhibitor (8905)	✓	✓
TT	DPP IV inhibitor (8911)	✓	✓
TV	DPP IV inhibitor (8912)	✓	
TVY	ACE inhibitor (7498; 15 μM)	✓	
TW	antioxidative (8459), DPP IV inhibitor (8913; 84 μM)	✓	
TY	antioxidative (8219), DPP IV inhibitor (8914)	✓	✓
VA	DPP IV inhibitor (3172; 168.24 μM)	✓	
VAF	ACE inhibitor (8126; 35.8 μM)	✓	
VAP	ACE inhibitor (3521; 2 μM)	✓	
VAV	ACE inhibitor (7635; 260 μM)	✓	
VAY	ACE inhibitor (3546; 16 μM)	✓	✓
VE	ACE inhibitor (7829), alpha-glucosidase inhibitor (9693; 22170 μM), DPP IV inhibitor (8916)	✓	
VF	ACE inhibitor (3384; 9.2 μM), DPP IV inhibitor (8917)	✓	✓
VFK	ACE inhibitor (7488; 1029 μM)	✓	
VG	ACE inhibitor (7594; 1100 μM), DPP IV inhibitor (8918)	✓	
VGL	DPP IV inhibitor (8694)	✓	
VH	DPP IV inhibitor (8919)	✓	
VI	DPP IV inhibitor (8920)	✓	✓
VIY	ACE inhibitor (7749; 7.5 μM)	✓	
VK	ACE inhibitor (7558; 13 μM), DPP IV inhibitor (8921)	✓	✓
VL	stimulating (8320), DPP IV inhibitor (8922; 74 μM)	✓	✓
VLY	ACE inhibitor (9050; 30.9 μM)	✓	
VM	DPP IV inhibitor (8923)	✓	
VPL	antiamnestic (3166; 47 μM), stimulating (3350), DPP IV inhibitor (8347; 15.8 μM)	✓	
VPW	antioxidative (8188)	✓	
VQ	DPP IV inhibitor (8925)	✓	
VR	ACE inhibitor (7628; 52.8 μM), DPP IV inhibitor (8594; 826.1 μM)	✓	✓
VRP	ACE inhibitor (3404; 2.2 μM)	✓	✓
VT	DPP IV inhibitor (8927)	✓	
VV	DPP IV inhibitor (3183)	✓	✓
VVF	ACE inhibitor (9044; 35.45 μM)	✓	
VVL	ACE inhibitor (9731)	✓	
VVV	anticancer (8318)	✓	
VW	ACE inhibitor (3486; 1.4 μM), antioxidative (8461), alpha-glucosidase inhibitor (9387; 22.6 μM), DPP IV inhibitor (8928)	✓	

VY	ACE inhibitor (3492; 7.1 μM), antioxidative (8224), DPP IV inhibitor (8929), DPP III inhibitor (9509)	✓	✓
WH	DPP IV inhibitor (8931)	✓	
WI	DPP IV inhibitor (8679; 138.7 μM)	✓	
WK	DPP IV inhibitor (8676; 40.6 μM)	✓	
WL	ACE inhibitor (9107; 41.4 μM), DPP IV inhibitor (8677; 43.6 μM)	✓	
WT	DPP IV inhibitor (8685; 482.1 μM)	✓	
YDY	antioxidative (7925)	✓	
YF	DPP IV inhibitor (8935), DPP III inhibitor (9480)	✓	
YH	ACE inhibitor (9087; 5.13 μM), DPP IV inhibitor (8937), DPP III inhibitor (9481)	✓	
YI	DPP IV inhibitor (8938), DPP III inhibitor (9510)	✓	✓
YK	ACE inhibitor (7697; 610 μM), DPP IV inhibitor (8939), DPP III inhibitor (9483)	✓	
YL	neuropeptide (8310), DPP IV inhibitor (8940), DPP III inhibitor (9482)	✓	✓
YLL	antioxidative (9349)	✓	✓
YM	DPP IV inhibitor (8941)	✓	
YN	ACE inhibitor (9185; 51 μM), DPP IV inhibitor (8942)	✓	
YQ	DPP IV inhibitor (8943)	✓	✓
YR	neuropeptide (9534), DPP IV inhibitor (8944), DPP III inhibitor (9484)	✓	✓
YS	DPP IV inhibitor (8945)	✓	
YT	DPP IV inhibitor (8696)	✓	
YV	ACE inhibitor (9077; 575.4 μM), DPP IV inhibitor (8946)	✓	
YVL	antibacterial (8268), antioxidative (8150)	✓	✓
YW	ACE inhibitor (3488; 10.5 μM), DPP IV inhibitor (8947)	✓	
YY	DPP IV inhibitor (8948), DPP III inhibitor (9476)	✓	

Leu or Ile could not be distinguished with the LC-MS/MS method used in this study. As Ile is slightly more hydrophilic than Leu, when two isomers with identical sequences except for one residue being either Leu or Ile were both found, the peptide eluted first on a C18 column was more likely to contain Ile (Lahrichi et al., 2013). When only one peak of the two isomers was found, or multiple Leu or Ile was found in a peptide sequence, the peptide identities usually could not be determined only based on the LC-MS/MS data. To further elucidate the identities of Leu and Ile in the peptides sequences, further studies using other strategies, such as applying higher

collision energy to produce diagnostic ions for Leu and Ile in LC-MS/MS analysis (Lahrichi et al., 2013) and comparing with synthetic peptides, will be needed.

Among the identified small bioactive peptides, 90 peptides were found in both the ENZ and CTRL protein-rich extracts. As the LC-MS/MS analysis performed in the present study was mainly for identification purposes, comprehensive relative quantification of the overlapping peptides was not performed. Nonetheless, it was still possible to know the approximate relative abundance by inspecting the MS1 peak areas. Most overlapping peptides had higher abundances in the ENZ sample than the CTRL sample because small peptides could be released from proteins and longer peptides where they were encrypted. For example, Val-Tyr is an ACE inhibitor that was shown to exhibit *in vivo* hypotensive activity previously. The peak area of nonderivatized Val-Tyr was substantially higher for the ENZ sample (4.8×10^6) compared to the CTRL sample (1.3×10^5), even though a higher volume of the CTRL sample was injected into the LC-MS/MS due to the relatively lower total peptide concentration (the injection volumes were 3 and 1 μL for the CTRL and ENZ samples, respectively, at the same dilution level).

Interestingly, Phe-Tyr, a dipeptide possessing both ACE inhibitory activity, which was demonstrated with a hypotensive effect in the animal model (Suetsuna et al., 2004), and antioxidative activity, was in high abundance in both the CTRL and ENZ protein-rich extracts. When injecting 3 and 1 μL of samples with the same dilution factor, the peak areas of nonderivatized Phe-Tyr were 1.4×10^7 and 1.5×10^7 for the CTRL and ENZ, receptively. It meant that Phe-Tyr might present in almond flour in a free form. Because Phe-Tyr was also found in several major proteins in almonds (e.g., prunin-1 (242-243) and (421-422), prunin-2 (195-196) and (337-338), and (R)-mandelonitrile lyase 2 (353-353) and (528-529)), it was likely that more

Phe-Tyr were released under the action of neutral protease during the extraction for the ENZ sample.

4.3.3.2. Identification of bioactivity for medium-sized peptides

In contrast to small peptides, although a significant number of medium-sized peptide sequences were found in the two protein-rich extracts (1739 peptide sequences in total), none of them matched with bioactive peptide sequences in the BIOPEP-UWM database. This might be due to the fact that studies on almond-derived bioactive peptides are still incomplete, or that the protein-rich extracts from almond flour contained medium-sized bioactive peptides that were not documented in the BIOPEP-UWM database. Because plants belonging to the same genus may share homologous protein sequences, they may generate identical bioactive peptide sequences during proteolysis. Therefore, we manually searched the literature for bioactive peptide sequences derived from *Prunus* genus seeds. A few angiotensin-converting enzyme (ACE) inhibitory peptides were found from protein hydrolysates generated from peach seeds and other *Prunus* genus fruit seeds in previous studies (González-García et al., 2018; Vásquez-Villanueva et al., 2019). One of those sequences, IYTPH, was also found in the ENZ protein-rich extract in the present study. Some ACE inhibitory peptides found in the fruit seed protein hydrolysates in the previous studies were also found encrypted in longer peptide sequences in the protein-rich extracts from almond flour (Table 4.2). For example, GIYSPH, a peptide identified from the ENZ protein-rich extract, contained the sequence IYSPH, which was shown exhibiting hypotensive activity previously (Vásquez-Villanueva et al., 2019). Nonetheless, additional amino acid residues may alter the ACE inhibitory activities. Understanding the contribution of additional residues on the activities and whether the bioactive sequences can be released during gastrointestinal digestion will need further investigation.

Table 4.2. Medium-sized Peptide sequences present in the protein-rich extracts from almond flour with ACE inhibitory peptide sequences encrypted.

Peptide sequence	Protein source	Active sequence	Identification	
			ENZ	CTRL
IYTPH[1]	Q43608 (Prunin-2)	IYTPH[2]	✓	
AIYTPH	Q43608 (Prunin-2)	IYTPH	✓	
GIYSPH	Q43607 (Prunin-1)	IYSPH	✓	
ADIFSPR	Q43607 (Prunin-1)	IFSPR[2]	✓	
IREGDVVAIPAV	Q43607 (Prunin-1)	VAIP[2]	✓	
IREGDVVAIPA	Q43607 (Prunin-1)	VAIP	✓	✓
RIREGDVVAIPA	Q43607 (Prunin-1)	VAIP	✓	
EGDVVAIPA	Q43607 (Prunin-1)	VAIP	✓	
REGDVVAIPA	Q43607 (Prunin-1)	VAIP	✓	
IREGDVVAIPAG	Q43607 (Prunin-1)	VAIP		✓
IREGDVVAIPAGVA	Q43607 (Prunin-1)	VAIP		✓
IREGDVVAIPAGVAYWS	Q43607 (Prunin-1)	VAIP		✓
TRRIREGDVVAIPAG	Q43607 (Prunin-1)	VAIP		✓
TRRIREGDVVAIPAGVA	Q43607 (Prunin-1)	VAIP		✓

[1] The peptide sequence marked in bold is identical to the previously reported active sequence.

[2] Vásquez-Villanueva et al., 2019.

4.4. Conclusions

The present is the first to apply dimethyl labeling as a tool to study small peptides and characterize the full peptidome of almond flour protein extracts. The substantial improvement in small peptide identification also demonstrated that the neutral protease, conventionally used in food processing, released a significant number of small peptides. Through further matching with the bioactive peptide database, more than 200 bioactive peptide sequences were identified. As most of the identified bioactive peptides were small peptides containing 2–4 amino acid residues, it was confirmed that peptidomic profiling of small peptides is essential for studying bioactive peptides in food protein hydrolysates. Not only this work revealed that almond proteins can be used as a substrate to generate peptides with various bioactivities, but the methods developed here

open a translational path for food byproduct valorization by investigating the peptidome in multiple types of food processing samples. Additionally, these methods will prove indispensable for unraveling the full complement of health-promoting peptides naturally found in fermented foods and, last but not least, to help explore the gastrointestinal digesta.

CRediT author contribution statement

Yu-Ping Huang: Conceptualization, Data curation, Investigation, Methodology, Visualization, Writing - original draft. **Fernanda Furlan Goncalves Dias:** Investigation, writing - review & editing. **Juliana Maria Leite Nobrega de Moura Bell:** Funding acquisition, Project administration, writing - review & editing. **Daniela Barile:** Conceptualization, Funding acquisition, Project administration, Writing - review & editing.

Acknowledgements

The authors thank Dr. Randall C. Robinson for providing advice in peptidomic analysis and Dr. Austin Horng-En Wang for providing technical support in bioactive peptide database matching. This study was supported by the U.S. Department of Agriculture's (USDA) Agricultural Marketing Service through grant AM170100XXXXG011 and the Hatch grant USDA: NIFA 2014-05266 and CA-D-FST-2187-H. The content of this publication is solely the responsibility of the authors and does not necessarily represent the official views of the USDA.

Supplementary material

Supplementary data can be downloaded at https://ars.els-cdn.com/content/image/1-s2.0-S0308814621018409-mmc1.xls, Table 4.S1. Small peptides identified in ENZ protein-rich extract from almond flour by LC-MS/MS analysis and de novo sequencing; Table 4.S2. Small peptides

identified in CTRL protein-rich extract from almond flour by LC-MS/MS analysis and de novo sequencing; Small peptides identified in CTRL protein-rich extract from almond flour by LC-MS/MS analysis and de novo sequencing; Table 4.S3. Medium-sized peptides identified in ENZ protein-rich extract from almond flour by LC-MS/MS analysis and database searches; Table 4.S4. Medium-sized peptides identified in CTRL protein-rich extract from almond flour by LC-MS/MS analysis and database searches; Table 4.S5. Major parent proteins of medium-sized peptides identified in the protein-rich extracts from almond flour.

References

Agyei, D., Tsopmo, A., & Udenigwe, C. C. (2018). Bioinformatics and peptidomics approaches to the discovery and analysis of food-derived bioactive peptides. *Analytical and Bioanalytical Chemistry*, *410*(15), 3463–3472. https://doi.org/10.1007/s00216-018-0974-1.

Bahna, S. L. (2008). Hypoallergenic formulas: Optimal choices for treatment versus prevention. *Annals of Allergy, Asthma & Immunology*, *101*(5), 453–459. https://doi.org/10.1016/S1081-1206(10)60281-5.

Costa, J., Mafra, I., Carrapatoso, I., & Oliveira, M. B. P. P. (2012). Almond allergens: Molecular characterization, detection, and clinical relevance. *Journal of Agricultural and Food Chemistry*, *60*, 1337–1349. https://doi.org/10.1021/jf2044923.

de Almeida, N. M., F. G. Dias, F., Rodrigues, M. I., & L. N. de Moura Bell, J. M. (2019). Effects of processing conditions on the simultaneous extraction and distribution of oil and protein from almond flour. *Processes*, *7*(11), 844. https://doi.org/10.3390/pr7110844.

de Souza, T. S. P., Dias, F. F. G., Koblitz, M. G. B., & de Moura Bell, J. M. L. N. (2020). Effects of enzymatic extraction of oil and protein from almond cake on the physicochemical and functional properties of protein extracts. *Food and Bioproducts Processing*, *122*, 280–290. https://doi.org/10.1016/j.fbp.2020.06.002.

Dias, F., de Almeida, N. M., S. P. de Souza, T., Taha, A. Y., & L. N. de Moura Bell, J. M. (2020). Characterization and demulsification of the oil-rich emulsion from the aqueous extraction process of almond flour. *Processes*, *8*(10), 1228. https://doi.org/10.3390/pr8101228.

Dias, J., Axelband, F., Lara, L. S., Muzi-Filho, H., & Vieyra, A. (2017). Is angiotensin-(3–4) (Val-Tyr), the shortest angiotensin II-derived peptide, opening new vistas on the renin–angiotensin system? *Journal of the Renin-Angiotensin-Aldosterone System*, *18*(1), 1470320316689338. https://doi.org/10.1177/1470320316689338.

Egger, L., Ménard, O., Delgado-Andrade, C., Alvito, P., Assunção, R., Balance, S., Barberá, R., Brodkorb, A., Cattenoz, T., Clemente, A., Comi, I., Dupont, D., Garcia-Llatas, G., Lagarda, M. J., Le Feunteun, S., JanssenDuijghuijsen, L., Karakaya, S., Lesmes, U., Mackie, A. R., … Portmann, R. (2016). The harmonized INFOGEST in vitro digestion method: From knowledge to action. *Food Research International*, *88*, 217–225. https://doi.org/10.1016/j.foodres.2015.12.006.

Eriksen, S. (1983). Application of enzymes in soy milk production to improve yield. *Journal of Food Science*, *48*(2), 445–447. https://doi.org/10.1111/j.1365-2621.1983.tb10762.x

FAOSTAT database. (2020). http://www.fao.org/faostat/en/#data/QI.

FoodData Central. (2020). https://ndb.nal.usda.gov/.

González-García, E., García, M. C., & Marina, M. L. (2018). Capillary liquid chromatography-ion trap-mass spectrometry methodology for the simultaneous quantification of four angiotensin-converting enzyme-inhibitory peptides in *Prunus* seed hydrolysates. *Journal of Chromatography A*, *1540*, 47–54. https://doi.org/10.1016/j.chroma.2018.02.003.

Gupta, R. S., Warren, C. M., Smith, B. M., Jiang, J., Blumenstock, J. A., Davis, M. M., Schleimer, R. P., & Nadeau, K. C. (2019). Prevalence and severity of food allergies among US adults. *JAMA Network Open*, *2*(1), e185630–e185630. https://doi.org/10.1001/jamanetworkopen.2018.5630.

Hong, S.-M., Tanaka, M., Koyanagi, R., Shen, W., & Matsui, T. (2016). Structural design of oligopeptides for intestinal transport model. *Journal of Agricultural and Food Chemistry*, *64*(10), 2072–2079. https://doi.org/10.1021/acs.jafc.6b00279.

Hsu, J.-L., Huang, S.-Y., Chow, N.-H., & Chen, S.-H. (2003). Stable-isotope dimethyl labeling for quantitative proteomics. *Analytical Chemistry*, *75*(24), 6843–6852. https://doi.org/10.1021/ac0348625.

Hsu, J.-L., Huang, S.-Y., Shiea, J.-T., Huang, W.-Y., & Chen, S.-H. (2005). Beyond quantitative proteomics: Signal enhancement of the a1 ion as a mass tag for peptide sequencing using dimethyl labeling. *Journal of Proteome Research*, *4*(1), 101–108. https://doi.org/10.1021/pr049837+.

Jin, Y., Yu, Y., Qi, Y., Wang, F., Yan, J., & Zou, H. (2016). Peptide profiling and the bioactivity character of yogurt in the simulated gastrointestinal digestion. *Journal of Proteomics*, *141*, 24–46. https://doi.org/10.1016/j.jprot.2016.04.010.

Koopman, R., Crombach, N., Gijsen, A. P., Walrand, S., Fauquant, J., Kies, A. K., Lemosquet, S., Saris, W. H., Boirie, Y., & van Loon, L. J. (2009). Ingestion of a protein hydrolysate is accompanied by an accelerated in vivo digestion and absorption rate when compared with its intact protein. *The American Journal of Clinical Nutrition*, *90*(1), 106–115. https://doi.org/10.3945/ajcn.2009.27474.

Korhonen, H., & Pihlanto, A. (2006). Bioactive peptides: Production and functionality. *International Dairy Journal*, *16*(9), 945–960. https://doi.org/10.1016/j.idairyj.2005.10.012.

Lacroix, I. M. E., & Li-Chan, E. C. Y. (2012). Dipeptidyl peptidase-IV inhibitory activity of dairy protein hydrolysates. *International Dairy Journal*, *25*(2), 97–102. https://doi.org/10.1016/j.idairyj.2012.01.003.

Lafarga, T., & Hayes, M. (2017). Bioactive protein hydrolysates in the functional food ingredient industry: Overcoming current challenges. *Food Reviews International*, *33*(3), 217–246. https://doi.org/10.1080/87559129.2016.1175013.

Lahrichi, S. L., Affolter, M., Zolezzi, I. S., & Panchaud, A. (2013). Food peptidomics: Large scale analysis of small bioactive peptides — a pilot study. *Journal of Proteomics*, *88*, 83–91. https://doi.org/10.1016/j.jprot.2013.02.018.

Li, S., Bu, T., Zheng, J., Liu, L., He, G., & Wu, J. (2019). Preparation, bioavailability, and mechanism of emerging activities of Ile-Pro-Pro and Val-Pro-Pro. *Comprehensive Reviews in Food Science and Food Safety*, *18*(4), 1097–1110. https://doi.org/10.1111/1541-4337.12457.

Liu, R.-L., Ge, X.-L., Gao, X.-Y., Zhan, H.-Y., Shi, T., Su, N., & Zhang, Z.-Q. (2016). Two angiotensin-converting enzyme-inhibitory peptides from almond protein and the protective action on vascular endothelial function. *Food & Function*, *7*(9), 3733–3739. https://doi.org/10.1039/C6FO00654J.

Locke, S. J., Leslie, A. D., Melanson, J. E., & Pinto, D. M. (2006). Deviation from the mobile proton model in amino-modified peptides: Implications for multiple reaction monitoring analysis of peptides. *Rapid Communications in Mass Spectrometry*, *20*(10), 1525–1530. https://doi.org/10.1002/rcm.2512.

Maestri, E., Marmiroli, M., & Marmiroli, N. (2016). Bioactive peptides in plant-derived foodstuffs. *Journal of Proteomics*, *147*, 140–155. https://doi.org/10.1016/j.jprot.2016.03.048.

Matsui, T., Li, C.-H., & Osajima, Y. (1999). Preparation and characterization of novel bioactive peptides responsible for angiotensin I-converting enzyme inhibition from wheat germ. *Journal of Peptide Science*, *5*(7), 289–297. https://doi.org/10.1002/(SICI)1099-1387(199907)5:7<289::AID-PSC196>3.0.CO;2-6.

McWilliam, V., Peters, R., Tang, M. L. K., Dharmage, S., Ponsonby, A.-L., Gurrin, L., Perrett, K., Koplin, J., Allen, K. J., Dwyer, T., Lowe, A., Wake, M., & Robertson, C. (2019). Patterns of tree nut sensitization and allergy in the first 6 years of life in a population-based cohort. *Journal of Allergy and Clinical Immunology*, *143*(2), 644-650.e5. https://doi.org/10.1016/j.jaci.2018.07.038.

Minkiewicz, P., Iwaniak, A., & Darewicz, M. (2019). BIOPEP-UWM database of bioactive peptides: Current opportunities. *International Journal of Molecular Sciences*, *20*(23), 5978. https://doi.org/10.3390/ijms20235978.

Murray, N. M., O'Riordan, D., Jacquier, J.-C., O'Sullivan, M., Holton, T. A., Wynne, K., Robinson, R. C., Barile, D., Nielsen, S. D., & Dallas, D. C. (2018). Peptidomic screening of bitter and nonbitter casein hydrolysate fractions for insulinogenic peptides. *Journal of Dairy Science*, *101*(4), 2826–2837. https://doi.org/10.3168/jds.2017-13853.

Nakamura, Y., Yamamoto, N., Sakai, K., Okubo, A., Yamazaki, S., & Takano, T. (1995). Purification and characterization of angiotensin I-converting enzyme inhibitors from sour milk. *Journal of Dairy Science*, *78*(4), 777–783. https://doi.org/10.3168/jds.S0022-0302(95)76689-9.

Rios-Villa, K. A., Bhattacharya, M., La, E. H., Barile, D., & Bornhorst, G. M. (2020). Interactions between whey proteins and cranberry juice after thermal or non-thermal processing during in vitro gastrointestinal digestion. *Food & Function*, *11*(9), 7661–7680. https://doi.org/10.1039/D0FO00177E.

Roberts, P. R., Burney, J. D., Black, K. W., & Zaloga, G. P. (1999). Effect of chain length on absorption of biologically active peptides from the gastrointestinal tract. *Digestion*, *60*(4), 332–337. https://doi.org/10.1159/000007679.

Sanjukta, S., & Rai, A. K. (2016). Production of bioactive peptides during soybean fermentation and their potential health benefits. *Trends in Food Science & Technology*, *50*, 1–10. https://doi.org/10.1016/j.tifs.2016.01.010.

Sasaki, M., Koplin, J. J., Dharmage, S. C., Field, M. J., Sawyer, S. M., McWilliam, V., Peters, R. L., Gurrin, L. C., Vuillermin, P. J., Douglass, J., Pezic, A., Brewerton, M., Tang, M. L. K., Patton, G. C., & Allen, K. J. (2018). Prevalence of clinic-defined food allergy in early adolescence: The SchoolNuts study. *Journal of Allergy and Clinical Immunology*, *141*(1), 391-398.e4. https://doi.org/10.1016/j.jaci.2017.05.041.

Suetsuna, K., Maekawa, K., & Chen, J.-R. (2004). Antihypertensive effects of *Undaria pinnatifida* (wakame) peptide on blood pressure in spontaneously hypertensive rats. *The Journal of Nutritional Biochemistry*, *15*(5), 267–272. https://doi.org/10.1016/j.jnutbio.2003.11.004.

Vásquez-Villanueva, R., Orellana, J. M., Marina, M. L., & García, M. C. (2019). Isolation and characterization of angiotensin converting enzyme inhibitory peptides from peach seed

hydrolysates: In vivo assessment of antihypertensive activity. *Journal of Agricultural and Food Chemistry*, *67*(37), 10313–10320. https://doi.org/10.1021/acs.jafc.9b02213.

Wang, X., Chen, H., Fu, X., Li, S., & Wei, J. (2017). A novel antioxidant and ACE inhibitory peptide from rice bran protein: Biochemical characterization and molecular docking study. *LWT*, *75*, 93–99. https://doi.org/10.1016/j.lwt.2016.08.047.

Wolf, W. J., & Sathe, S. K. (1998). Ultracentrifugal and polyacrylamide gel electrophoretic studies of extractability and stability of almond meal proteins. *Journal of the Science of Food and Agriculture*, *78*(4), 511–521. https://doi.org/10.1002/(SICI)1097-0010(199812)78:4<511::AID-JSFA148>3.0.CO;2-X.

Zhang, J., Xin, L., Shan, B., Chen, W., Xie, M., Yuen, D., Zhang, W., Zhang, Z., Lajoie, G. A., & Ma, B. (2012). PEAKS DB: De novo sequencing assisted database search for sensitive and accurate peptide identification. *Molecular & Cellular Proteomics*, *11*(4). https://doi.org/10.1074/mcp.M111.010587.

Chapter V

Comprehensive oligosaccharide profiling of commercial almond milk, soy milk, and soy flour

(This chapter is a manuscript under peer review; authors: **Huang, Y.-P.**, Paviani, B., Fukagawa, N.K., Phillips, K.M., & Barile, D.)

Abstract

Almond milk, soy milk, and soy flour are dietary sources of oligosaccharides that could beneficially affect human health but in sensitive individuals cause intestinal discomfort. This study quantified the oligosaccharides raffinose, stachyose, and verbascose in commercial products by high-performance anion-exchange chromatography-pulsed amperometric detection (HPAE-PAD). The extraction and quantification methods were optimized and validated to assure measurement accuracy and repeatability (91–107% recovery and 0.0–5.4% intra- and inter-day RSD). The summed concentration of raffinose, stachyose, and verbascose was in the range of 0.118–0.19 mg/g in almond milk, 3.6–6.4 mg/g in soy milk, 74–77 mg/g in defatted soy milk, and 4.8–57 mg/g in full-fat soy four. A comprehensive oligosaccharide profiling was conducted with liquid chromatography-tandem mass spectrometry (LC-MS/MS). Over 80 oligosaccharides with various structures (e.g., $Hex_{3–8}$, ciceritol, and $Hex_{2–3}Glycerol_1$) were overall identified. Additionally, novel compounds, 2,3-butanediol glycosides, were identified in almond milk in significant abundance.

Keywords: α-galactooligosaccharides; HPAE-PAD; LC-Q-TOF MS; 2,3-butanediol glycosides; low molecular weight soluble dietary fiber; soybeans

5.1. Introduction

Non-digestible oligosaccharides (NDO) are carbohydrates with intermediate sizes between simple sugars and polysaccharides resistant to hydrolysis in the upper gastrointestinal tract. Naturally occurring NDO can be found in various foods with diverse structures. Depending on the structures, NDO may be utilized by beneficial gut bacteria, such as *Bifidobacterium* and *Lactobacillus*, to promote their growth and activity. The selective stimulation of the growth and

activity of beneficial bacteria is associated with multiple potential advantages to human health, such as reducing the risk of colorectal cancer, modulating the immune system, improving mineral absorption, and regulating lipid metabolism (Swennen, Courtin, & Delcour, 2006). Because of the potential health benefits, NDO are also produced through different techniques (e.g., extraction and enzymatic treatment) as functional food ingredients or dietary supplements.

Consumers' preference to include more plant-based foods in the diet has been surging in recent years. The trend is related to consumers' perception that plant-based foods are usually healthier and have a lower environmental impact than animal-based foods. This significant shift in consumers' attitudes toward plant-based foods resulted in increased development and consumption of plant-based beverages, such as almond milk, soy milk, and others, as alternatives to cow's milk. Soy milk is a traditional beverage widely consumed in several Asian countries, such as China, Thailand, and Taiwan, and has only recently become more prevalent in Western countries. Almond milk, followed by soy milk, is currently the most popular milk alternatives in the United States (Wunsch, 2022). These plant-based beverages also serve as alternative options for consumers who are allergic to cow's milk. Plant-based food ingredients are also increasingly used to produce foods for consumers with special dietary restrictions or can be added to fortify specific nutrients. For example, soy flour is a common ingredient used to replace wheat flour in producing gluten-free foods. Furthermore, it is often combined with cereals to supplement the essential amino acids lacking in cereals (Sereewat et al., 2015).

Soybean and almond seeds both contain naturally occurring oligosaccharides, predominated by α-galactooligosaccharides which include a sucrose core (Glu-α-1,β-2-Fru) extended with one or more galactose residues with α-1,6-glycosidic linkages. Raffinose, stachyose, and verbascose (degree of polymerization = 3, 4, and 5, respectively) are the major

oligosaccharides in soybean and almonds. Due to the abundance of α-galactooligosaccharides in soybeans (32–43 mg/g stachyose, 8–13 mg/g raffinose, and 1–2 mg/g verbascose) (Fan, Zang, & Xing, 2015; Kuo, VanMiddlesworth, & Wolf, 1988), the consumption of soybean products may cause flatulence (Liener, 1994). The flatulence-causing potential engendered the perception of α-galactooligosaccharides as undesirable components and urged numerous studies seeking approaches to remove them from soybean and other legume products (Liener, 1994). Nonetheless, as mentioned above, the gas-producing property is related to microbial fermentation, which could benefit human health. Indeed, α-galactooligosaccharides have been shown to exhibit potential prebiotic activity, such as reshaping bacterial composition by increasing the relative abundance of *Bifidobacterium* and *Lactobacillus* and in turn producing short-chain fatty acids in *in vitro* and rodent models (Amorim et al., 2020; Xi et al., 2021). Consumption of α-galactooligosaccharides could lead to additional beneficial health effects, including decreasing total cholesterol and low-density lipoprotein cholesterol, as shown in mice fed with a high-fat diet (Chappuis, Morel-Depeisse, Bariohay, & Roux, 2017; Dai et al., 2019) and reducing appetite and inflammation in overweight adults (Morel et al., 2015).

Although a few investigations report the contents of some oligosaccharides in almonds and soybean (Barreira, Pereira, Oliveira, & Ferreira, 2010; Fan et al., 2015; Kuo et al., 1988), those values cannot be used for extrapolating the oligosaccharides concentration in commercial almond and soy products. Oligosaccharide contents in plants may vary with varieties, geography, and growing conditions; the contents in raw materials can also be significantly altered during processing, such as fractionation and thermal treatments (Bainy, Tosh, Corredig, Poysa, & Woodrow, 2008). The processing procedure and formulation of almond and soy milk often differ among manufacturers, resulting in varying oligosaccharide concentrations in commercial

products. Therefore, there is a need for a well-designed study of various commercial products to survey their oligosaccharide concentrations and assess the levels in commonly found products.

Besides α-galactooligosaccharides, other oligosaccharides present in plant-based foods may be involved in gut microbial fermentation and should be included in the characterization. Although pure standards are only available for a limited subset of well-studied oligosaccharides, liquid chromatography-tandem mass spectrometry (LC-MS/MS) can be used to comprehensively characterize all oligosaccharides through deducing structural information based on unique fragmentation patterns. Mass spectrometric oligosaccharide profiling has been applied to human milk, bovine milk, and goat milk (Aldredge et al., 2013; Lu et al., 2020; Wu, Grimm, German, & Lebrilla, 2011) but is still underutilized on plant-based foods to date. Recently, our group reported the optimization of a conventional LC-MS/MS workflow to improve the identification of low-abundant oligosaccharides (Huang, Robinson, Dias, de Moura Bell, & Barile, 2022) and avoid incorrect oligosaccharide identification caused by unexpected oligosaccharide degradation (Huang, Robinson, & Barile, 2022), enabling its application to the identification of plant oligosaccharides.

The current study aimed to profile the oligosaccharides in two major types of plant-based beverages (almond milk and soy milk) and in soy flour. Quantification of major oligosaccharides, including raffinose, stachyose, and verbascose, was performed by high-performance anion-exchange chromatography with pulsed amperometric detection (HPAE-PAD) using pure standards. To ensure the quantification method's reliability and ease, oligosaccharide extraction prior to HPAE-PAD analysis was optimized, and a method validation was carried out. A comprehensive oligosaccharide identification was achieved by liquid chromatography-

quadrupole-time-of-flight mass spectrometry (LC-Q-TOF MS) to profile the unexplored minor oligosaccharides. Additionally, the structures of selected unknown glycosides were elucidated.

5.2. Materials and methods

5.2.1. Almond milk, soy milk, and soy flour

The current dataset included four types of commercial products, including unsweetened almond milk (eight brands, AM1–AM8), unsweetened soy milk (eight brands, SM1–SM8), defatted soy flour (five brands, DFSF1–DFSF5), and full-fat soy flour (six brands, FFSF1–FFSF6), produced by different food manufacturers (Supplementary material Table 5.S1); ingredients of the almond milk and soy milk products varied with brands. Samples analyzed in this work are those reported in the U.S. Department of Agriculture FoodData Central (https://fdc.nal.usda.gov), where information about other food components may be found (FoodData Central, 2022). Two additional soy flour samples, NIST SRM® 3234 (National Institute of Standards and Technology (Gaithersburg, MD) and in-house full fat soy flour control material ("Soy Flour CC"), were analyzed in each batch for quality control.

5.2.2. Reagents

Raffinose (product 95068, purity 99.0%), stachyose (product S4001, purity 98%), verbascose (product 56217, purity 97.3%), trifluoroacetic acid (TFA), 2,3-butanediol (product 42038), and trichloroacetyl isocyanate were purchased from MilliporeSigma (St. Louis, MO, USA). n-Hexane, Carrez solutions I and II, acetonitrile (LC-MS grade), and formic acid (LC-MS grade) were obtained from Fisher Scientific (Waltham, MA, USA). Water was obtained from a Direct-Q 5 UV water purification system (18.2 MΩ cm at 25 °C) (EMD Millipore, now part of MilliporeSigma).

5.2.3. Soy flour defatting

n-Hexane was added to 20–25 mg of full-fat soy flour samples (20:1, v/w) in 1.5 mL tubes. The samples were vortexed to suspend the soy flour and shaken on a thermomixer (ThermoMixer C, Eppendorf, Hamburg, Germany) at 1,400 rpm at room temperature for 10 min. After the samples were centrifuged at 13,000 g, the supernatants were removed. The soy flour pellets were re-extracted with n-hexane (20:1, v/w) under the same condition. The supernatant was discarded after centrifuge. The defatted soy flour pellets were dried with a centrifugal evaporator (Genevac miVac concentrator, SP Scientific, Warminster, PA, USA) to remove residual n-hexane.

5.2.4. Quantification of raffinose, stachyose, and verbascose

5.2.4.1. Method development for extraction of oligosaccharides from almond milk and soy milk

5.2.4.1.1. Carrez clarification

Almond milk and soy milk samples (200 μL) were transferred to 1.5 mL tubes, diluted with water, and added with 15, 20, 25, 50, and 100 μL of Carrez I solution (85 mM $K_4[Fe(CN)_6]$) and an equal volume of Carrez II solution (250 mM $ZnSO_4$). The total liquid volume of each sample was 1 mL. The samples were vortexed and shaken at 1,000 rpm for 5 min on a thermomixer. After being centrifuged at 13,000 g at room temperature for 15 min, the supernatants were transferred to new 1.5 mL tubes. The efficacy of clarification was evaluated by eyes.

The clarification was re-conducted using the Carrez solution volume with the best clarification efficacy. The supernatants were transferred to 2 mL volumetric flasks after clarification and centrifuge. One milliliter of water was added to the pellets to suspend the pellets with pipette tips. After the samples were shaken at 1,000 rpm at room temperature for 5 min and centrifuged at 13,000 g at room temperature for 15 min, the supernatants were combined with the

first supernatants in the volumetric flasks. Additional water was added to the mark to bring the final volume to 2 mL.

5.2.4.1.2. Ethanol precipitation

Protein precipitation using two (2V) and four (4V) volumes of ethanol was tested. For the 2V samples, almond milk and soy milk samples (200 μL) were diluted by adding 133 μL of water. Cold ethanol (667 and 800 μL) was added to the diluted samples and 200 μL of undiluted samples, respectively, for the 2V and 4V samples. The samples were vortexed and incubated at −20 °C for 1 h. After being centrifuged at 13,000 g at 4°C for 30 min, the supernatants were transferred to new tubes. One milliliter of 66.7% and 80% cold ethanol (v/v) was added to the tubes with pellets for the 2V and 4V samples, respectively. The pellets were dispersed with the aid of pipette tips. After shaking the samples at 1,000 rpm at room temperature for 5 min, the samples were centrifuged at 13,000 g at 4 °C for 30 min. The two supernatants were combined and dried with a centrifugal evaporator. The dried samples were dissolved in water and diluted to 2 mL using volumetric flasks.

5.2.4.2. Extraction method development for soy flour

Oligosaccharides in soy flour were extracted with water with simultaneous or separate Carrez clarification. For simultaneous extraction and clarification, 25 mg of soy flour was extracted with 900 μL of water by shaking at 1,500 rpm at room temperature for 10 min. Carrez I and Carrez II solutions (50 μL each) were subsequently added to the soy flour-water mixture. After being shaken at 1,500 rpm at room temperature for 5 min and centrifuged at 13,000 g at room temperature for 15 min, the supernatants were transferred to 2 mL volumetric flasks. One milliliter of water was added to the pellets to suspend them. The samples were shaken at 1,500 rpm at room

temperature for 5 min. After centrifuge, the supernatants were combined with the previous extracts in the volumetric flasks. The total volume was brought to 2 mL by adding water.

For separate extraction and clarification, 25 mg of soy flour was also extracted with 900 μL of water by shaking at 1,500 rpm at room temperature for 10 min. The samples were centrifuged at 13,000 g at room temperature for 15 min. After transferring the supernatants to new tubes, 900 μL of water was added to the soy flour pellets to extract residual oligosaccharides by shaking at 1,500 rpm at room temperature for 5 min. After centrifuging at 13,000 g at room temperature for 15 min, the supernatants were combined with the previous extract. Carrez I and Carrez II solutions (50 or 100 μL each) were added to the combined supernatants. The samples were shaken at 1,500 rpm at room temperature for 5 min. After centrifuge, the supernatants were transferred to 5 mL volumetric flasks. The pellets were suspended using pipette tips after adding 1 mL of water. The samples were shaken at 1,500 rpm at room temperature for 5 min and centrifuged at 13,000 g at room temperature for 15 min. The supernatants were combined with the previous supernatants in the volumetric flasks. Extra water was added to the mark to make a final sample volume of 5 mL.

5.2.4.3. High-performance anion-exchange chromatography with pulsed amperometric detection

Quantification of raffinose, stachyose, and verbascose was conducted on a ThermoFisher Dionex ICS-5000+ high-performance anion-exchange chromatography with pulsed amperometric detection (HPAE-PAD) system equipped with a CarboPac PA200 guard column (3 × 50 mm) and a CarboPac PA200 analytical column (3 × 250 mm). Chromatographic separation was carried out with a 15-min isocratic elution using 40 mM NaOH at a flow rate of 0.5 mL/min. The column was flushed with 200 mM NaOH for 5 min at the end of each run and equilibrated with 40% NaOH for 10 min before the next injection. Calibration curves were constructed by injecting standard solutions with concentrations of 0.1−10 μg/mL. The concentration of raffinose, stachyose, and

verbascose in samples was calculated with the calibration curves. The regression model (linear versus quadratic) for the calibration curves was selected by comparing the two models using a partial F-test (Massart et al., 1998) with R programming (version 3.5.3).

5.2.4.4. Quantification method validation

The quantification method was validated with the instrumental limit of detection (LOD) and limit of quantification (LOQ), coefficient of determination (r^2), recovery, and intra- and inter-batch relative standard deviation (RSD) of quality control samples. LOD and LOQ were determined with the signal of noise ratios (S/N) of 3 and 10, respectively.

5.2.4.5. Recovery

Because it is not possible to find "blank" almond milk, soy milk, and soy flour samples that are free of raffinose, stachyose, and verbascose, known amounts of oligosaccharides were spiked to the almond and soy flour samples to measure the recovery by subtracting oligosaccharide concentration in the unspiked samples from the spiked samples. After spiking raffinose, stachyose, and verbascose standards (~25−50% of the original amounts of each oligosaccharide in the samples; the original amounts of oligosaccharides were calculated with the concentrations determined in a preliminary experiment) to almond milk, soy milk, and soy flour samples, the spiked and unspiked samples were extracted with the selected procedures and analyzed by HPAE-PAD. The recovery was calculated by dividing the differences between the spiked and unspiked samples by the spiked amounts. The recovery represented the extraction recovery and the instrument measurement recovery.

5.2.4.6. Batch extraction with the selected procedures

The samples of commercial almond milk, soy milk, defatted soy flour, and full-fat soy flour were arranged into three assay batches (Table 5.S1). Samples in the same assay batch were

extracted and analyzed on the instrument together. Each sample was extracted and analyzed once (n = 1) except that selected samples, as specified in Table 5.S1, were extracted and analyzed multiple times (n = 2–4) to estimate the overall measurement uncertainty.

Almond milk and soy milk samples (200 μL, weights recorded) were diluted to 960 and 900 μL, respectively, with water. Carrez I and Carrez II solutions (20 and 50 μL, respectively, each) were then added to make a total volume of 1 mL. After vortexed, the tubes were shaken at 1,400 rpm at room temperature for 5 min. The samples were centrifuged at 13,000 g at room temperature for 15 min. The supernatants were transferred to 2 mL volumetric flasks. One milliliter of water was slowly added to the pellets; the pellets were suspended with the pipette tips in the meantime. After being vortexed, the tubes were shaken (at 1,400 rpm and room temperature for 5 min) and centrifuged (at 13,000 g and room temperature for 15 min). The supernatants were transferred to the corresponding volumetric flasks and combined with the first extract. The total volume of each sample was brought to 2 mL by adding water. The samples were passed through 0.2 μm filters and analyzed by HPAE-PAD.

5.2.5. Comprehensive identification of oligosaccharides

5.2.5.1. Oligosaccharide purification

Almond milk (8 samples, AM1–AM8), soy milk (8 samples SM1–SM8), defatted (5 samples, DFSF1–DFSF5), and full-fat (6 samples, FFSF1–FFSF6) soy flour samples were respectively pooled by combining equal amounts of each sample. The pooled almond milk and soy milk samples were mixed with 4 volumes of cold ethanol. After incubating at -30 °C for 1 h, the samples were centrifuged at 13,000 g at 4 °C for 30 min. The supernatants were transferred to new tubes and dried in a centrifugal evaporator. The dried samples were dissolved in 0.1 % formic acid in water (v/v) and purified by two solid-phase extraction (SPE) steps. Mixed-mode SPE was

applied to remove hydrophobic interferences, such as proteins and peptides, and improve oligosaccharide identification as described previously (Huang, Robinson, Dias, et al., 2022) with slight modification. Briefly, mixed-mode SPE cartridges (Strata-X-C, 30 mg; Phenomenex, Torrance, CA, USA) were activated with 2 mL of acetonitrile and equilibrated with 2 mL of 0.1% formic acid in water. The re-dissolved samples were loaded onto the cartridges, which were then flushed with 3 mL of 0.1% formic acid in water. The eluates during sample loading and further flushing with 0.1% formic acid were collected in one tube. The oligosaccharide samples collected from the mixed-mode SPE were further purified with a porous graphitic carbon (PGC) SPE microplate with 40 μL chromatographic media bed (Glygen Corporation, Columbia, MD, USA). The PGC microplate was conditioned with 300 μL of 80% acetonitrile, 0.1% TFA in water (v/v/v) and 300 μL of water. For sample loading, 600 μL of the oligosaccharide fractions were added to the conditioned microplate at a time until the whole samples were loaded. The microplate wells were washed with 400 μL of water for three times. The oligosaccharides were recovered in two fractions. Neutral (F1) and acidic (F2) oligosaccharides were eluted with 40% acetonitrile in water (v/v) and 40% acetonitrile, 0.1% TFA in water (v/v/v), respectively. The fractions were dried in a centrifugal evaporator at room temperature.

5.2.5.2. LC-Q-TOF MS analysis

The purified oligosaccharides were analyzed by an Agilent 6520 Accurate-Mass Q-TOF LC-MS with a Chip Cube interface (Agilent Technologies, Santa Clara, CA, USA). The mobile phase was composed of 5 mM ammonium acetate, 3% acetonitrile in water (solvent A) and 5 mM ammonium acetate, 90% acetonitrile in water (solvent B). The ammonium salt in the mobile phase can promote the formation of ammonium species, which can aid the differentiation of authentic oligosaccharides and in-source fragments as well as avoid incorrect identification (Huang,

Robinson, & Barile, 2022). The oligosaccharide samples were delivered to the enrichment column of an Agilent PGC-Chip II (G4240-64010) with 100% A at a flow rate of 4 μL/min. The oligosaccharides were separated on the analytical column of the PGC chip with a 60-min gradient. The gradient started from 100% A, increased from 0 % to 16 % B in 20 min, from 16% to 44% B in 10 min, from 44 to 100% B in 5 min, and was held at 100% B for 10 min. The system was equilibrated at 100% A for 15 min before 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 1875 V. The ions were scanned within the range of *m/z* 150–2500 at a rate of 1 spectrum sec^{-1}. 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). The active exclusion was enabled. Reference ions *m/z* 922.009798 and *m/z* 1221.990637 were used for continual mass calibration throughout the analysis.

2.5.3. LC-Q-TOF MS data analysis

Oligosaccharides were identified by manually inspecting fragmentation patterns in tandem MS spectra. Peak area integration was fulfilled with Profinder B.08.00 (Agilent Technologies). Targeted feature extraction was conducted based on a library including the monoisotopic masses and retention times for all the oligosaccharide identifications in each type of sample. Signals of $[M + H]^+$, $[M + Na]^+$, $[M + K]^+$, and $[M + NH_4]^+$ with a mass error within 20 ppm were included for peak area integration. For raffinose, stachyose, and verbascose, in-source fragment ions as well as dimer and trimer aggregates, were also included to approach their actual abundance (Huang, Robinson, & Barile, 2022). The apparent relative abundance of each oligosaccharide was calculated with the peak area of individual oligosaccharides divided by the total peak area of all the identified oligosaccharides.

5.2.6. Identification of selected unknown glycosides

5.2.6.1. Glycosidase treatment

To identify the unknown glycosides present in almond milk, glycosidases were applied to the purified almond oligosaccharide sample. Five microliters of the purified almond neutral oligosaccharide (F1) sample were mixed with 0.1 unit of α-glucosidase (G0660, MilliporeSigma), 0.1 unit of β-glucosidase (G4511, MilliporeSigma), or water (as a control) and incubated at 37 °C in a thermomixer with shaking at 700 rpm for 30 min. After glycosidase treatment, the samples were cleaned up with C18 SPE and PGC SPE. C18 SPE was conducted with a microplate with 40 μL chromatographic media bed (Glygen Corporation). The samples were loaded to the microplate wells preconditioned with 300 μL of acetonitrile and 300 μL of water. The oligosaccharides were eluted with 600 μL of water and further loaded to PGC SPE microplate wells with 40 μL chromatographic media bed (Glygen Corporation) preconditioned with 300 μL of 80% acetonitrile (v/v) followed by 300 μL of water. The oligosaccharides were collected during the initial sample loading and the subsequent elution with 600 μL of 40% acetonitrile in water (v/v). The samples were dried with a centrifugal evaporator at room temperature. After dissolving in water, the glycosidase-treated samples were analyzed by the LC-Q-TOF MS as described above.

5.2.6.2. Analysis of the aglycone using liquid chromatography-triple quadrupole mass spectrometry (LC-QqQ MS) analysis

The aglycone was tentatively identified as 2,3-butanediol based on its *m/z* value. Due to the poor ionization in the native form, 2,3-butanediol was analyzed after derivatization with trichloroacetyl isocyanate as described by Chen et al. (2018) with some modifications. The β-glucosidase treated almond oligosaccharide samples (5 μL) were diluted 100 times with acetonitrile, and 5 μL of trichloroacetyl isocyanate were added to the samples. After vortexing for

1 min, the samples were dried in a centrifugal evaporator. The dried samples were dissolved in 50% acetonitrile in water and injected into an Agilent 6470 Triple Quadrupole Liquid Chromatography-Mass Spectrometry System equipped with a Zorbax Eclipse Plus C18 column (2.1 × 50 mm, 1.8 μm, Agilent). The mobile phase consisted of 10 mM ammonium acetate in 3% water, 97% acetonitrile (pH 4.5; A) and 10 mM ammonium acetate in 95% acetonitrile, 5% water (pH 4.5; B). The chromatographic separation was carried out at 40 °C with gradient elution at a flow rate of 0.6 mL/min starting from 15% B. The eluent was kept at 15% B from 0−2 min and increased to 45% B from 2−14 min. After the LC separation, the column was regenerated by flushing with 100% B for 2 min and equilibrated at 15% B for 4 min before the next injection. The MS analysis was conducted in positive ion mode with source parameters as follows: the gas temperature was 200 °C at a flow rate of 11 L/min; the nebulizer was 35 psi; the sheath gas temperature was 200 °C at a flow rate of 10 L/min; capillary voltage was 3000 V. Multiple reaction monitoring (MRM) of two transitions, *m/z* 484 → 260 and *m/z* 484 → 262, was conducted with fragmentor of 135 V and collision energy of 15 V for both transitions from 5 to 14 min.

5.3. Results and discussion

5.3.1. Quantification of raffinose, stachyose, and verbascose by HPAE-PAD

5.3.1.1. HPAE-PAD method

An HPAE-PAD method using a CarboPac PA200 column with isocratic elution was developed to separate and quantify major oligosaccharides in the almond and soy samples. Fig. 5.1 shows the chromatogram of a mixture of raffinose, stachyose, and verbascose standards. Although stachyose was eluted immediately after raffinose, an ideal resolution (2.25–2.66; calculated using the peak widths at 50% height) between the two peaks was assured.

Linear regression models are often used for creating calibration curves in carbohydrate quantification using HPAE-PAD (Pico, Martínez, Martín, & Gómez, 2015; Pico et al., 2021), but quadratic models may result in a better fit and higher accuracies in some cases (Haselberger & Jacobs, 2016; Ispiryan, Heitmann, Hoehnel, Zannini, & Arendt, 2019). In order to determine which model was more suitable, we compared the linear and quadratic models with a partial F-test (Massart et al., 1998). The test results showed that the addition of the quadratic term significantly improved the model ($p < 0.001$) for the calibration curves of all the three oligosaccharides analyzed. Therefore, quadratic calibration curves were chosen to determine the oligosaccharide concentrations.

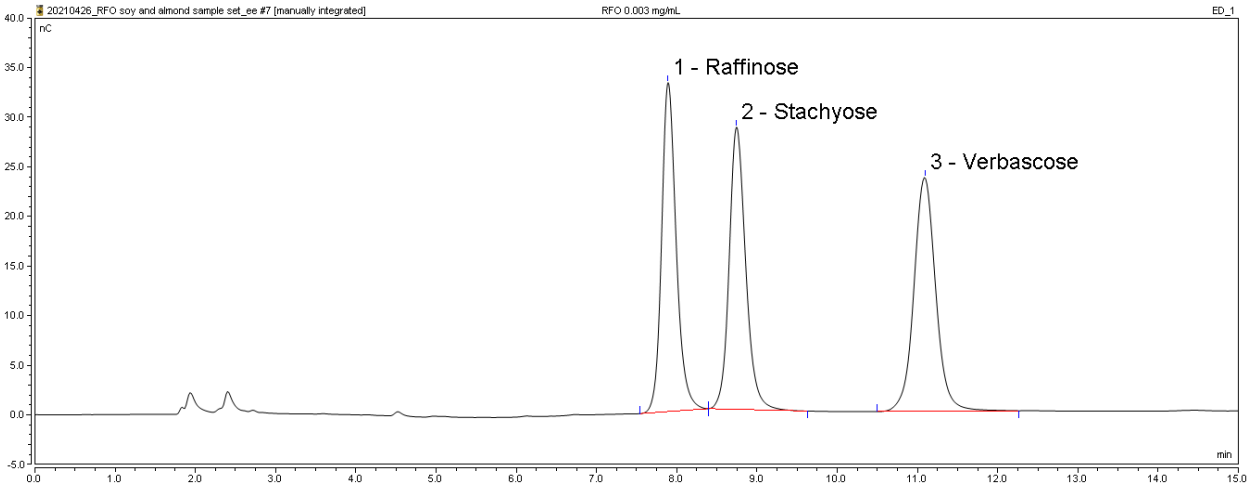


Fig. 5.1. HPAE-PAD chromatogram of standards of raffinose, stachyose, and verbascose.

5.3.1.2. Extraction method development

To ensure quantification accuracy, we compared several oligosaccharide extraction procedures for the liquid (almond milk and soy milk) and solid (soy flour) samples. After selecting appropriate procedures, the recovery was determined by measuring the known amount of oligosaccharides spiked in the samples.

5.3.1.3. Extraction of almond milk and soy milk

Almond milk and soy milk contain varied protein content (0.44−0.69 g/100g and 3−4.69 g/100g, respectively) (FoodData Central, 2022). Hydrocolloids, such as gellan gum, carrageenan, and locust bean gum, are often used in commercial plant-based beverages as thickeners and stabilizers. These components need to be eliminated prior to HPAE-PAD analysis to protect the chromatographic column and ensure analysis quality. Ethanol precipitation and Carrez clarification are often used to remove these large molecules and thus were both tested in this study. Different levels of Carrez solution were tested for the clarification of almond milk and soy milk because the efficacy of clarification would be associated with the sample composition. When using 30−200 μL of Carrez (Carrez I + Carrez II) solution to clarify almond milk, the supernatants were completely clear for all the five levels tested. In comparison, only 100 and 200 μL of Carrez solution resulted in completely clear supernatants for the clarification of soy milk. The turbidity of the supernatants from the samples clarified with 30, 40, and 50 μL of Carrez solution decreased as the Carrez solution volume increased. Based on these results, clarification using 50 and 100 μL of Carrez solutions for almond milk and soy milk, respectively, was further evaluated by HPAE-PAD analysis and compared with ethanol precipitation.

The efficiency of oligosaccharide extraction from almond milk and soy milk by ethanol precipitation and Carrez clarification was evaluated by comparing the HPAE-PAD quantification values for raffinose, stachyose, and verbascose as shown in Table 5.S2 For almond milk, the measured quantities of the oligosaccharides were similar among the three procedures, including ethanol precipitation using 2 and 4 volumes of ethanol and Carrez precipitation, except that stachyose was slightly lower for precipitation with 4 volumes of ethanol than the other two procedures. For soy milk, ethanol precipitation using 2 volumes of ethanol and Carrez precipitation

resulted in similar oligosaccharide quantities. However, precipitation using 4 volumes of ethanol led to significantly lower measurement values for all the three oligosaccharides. Bouchard, Hofland, and Witkamp (2007) reported that the solubility of raffinose at 310 K was higher in water than in water-ethanol mixtures; the solubility also decreased as ethanol percentages in water-ethanol mixtures increased. Therefore, the lower extraction efficiency of 4 volumes of ethanol in the current study might be attributed to the lower solubility of oligosaccharides in 80% ethanol than in 66.7% ethanol.

All the three extraction procedures tested in this study resulted in clean chromatogram background. The samples also did not cause back pressure increase in HPAE-PAD analysis. Thus, the efficacy of large molecule removal by using the three procedures was considered sufficient. When using Carrez clarification, the samples can be directly injected into HPAE-PAD, usually on the same day, after filtration and appropriate dilution. In contrast, ethanol needs to be evaporated, and samples must be re-dissolved in water before HPAE-PAD analysis, leading to an extended analysis time. Consequently, Carrez clarification was selected for the batch extraction. The recovery of raffinose, stachyose, and verbascose was checked by measuring the spiked and unspiked samples and ranged from 91−107% (Table 5.1).

Table 5.1. Instrumental limit of detection (LOD) and quantification (LOQ), recovery, relative standard deviation (RSD), coefficients of determination (r^2) of quadratic calibration curves of the quantification method used for batch extraction.

Oligosaccharide	LOD (μg L^{-1})	LOQ (μg L^{-1})	Recovery (%)			RSD (%)[2]		r^2 [3]
			AM[1]	SM	SF	Intra-batch	Inter-batch	
Raffinose	1.2	4.9	107 ± 3	101 ± 3	103 ± 2	0.0−2.4	1.5−5.4	0.999991−0.999995
Stachyose	1.3	6.3	93 ± 3	91 ± 4	101 ± 5	0.2−2.6	1.3−5.2	0.999997−0.999999
Verbascose	3.4	12	100 ± 2	103 ± 4	103 ± 0	0.1−3.8	2.7−5.2	0.999992−0.999998

[1]AM: almond milk; SM: soy milk; SF: soy flour (defatted).

[2] Intra-batch RSD was obtained from duplicate analysis on one or two selected samples in each batch; inter-batch RSD was measured by analyzing two quality control samples in each batch.

[3] The r^2 values represent the range among all batches.

5.3.1.4. Extraction of soy flour

For the extraction of soy flour, we selected water as the solvent and used Carrez clarification to remove water-soluble proteins for the following reasons. As mentioned above, others (Bouchard et al., 2007; Pico et al., 2015) found that the solubility of small carbohydrate molecules was higher in pure water than in water-ethanol mixtures, in agreement with our observation that the extraction efficiency of 80% ethanol was lower than 66.7% ethanol for soy milk oligosaccharides. Moreover, it would be best to use consistent extraction techniques for the "milk" samples and soy flour. The quantification values obtained from the samples prepared by aqueous extraction followed by either separate or combined Carrez clarification were shown in Table 5.S3 For separate Carrez clarification, using 100 and 200 μL of Carrez solution resulted in similar oligosaccharide quantification values. For combined Carrez clarification, the oligosaccharide quantification values had no significant difference from separate Carrez clarification. Because using combined Carrez clarification requires fewer steps for the extraction procedure, it was selected to be used for the batch extraction of soy flour. Satisfying recovery of the method, measured by spiking known amounts of oligosaccharides, was achieved (100, 103, and 103% for raffinose, stachyose, and verbascose, respectively).

5.3.1.5. Quantification of raffinose, stachyose, and verbascose in commercial products

Fig. 5.2 shows the raffinose, stachyose, and verbascose concentrations in different brands of almond milk, soy milk, and soy flour. The concentrations in the beverages were measured in the unit on a weight-to-weight basis (mg/g) to avoid pipetting inaccuracy due to the affinity of

proteins to pipette tips. Almond milk contained 0.056−0.11 mg/g of raffinose, 0.046−0.085 mg/g of stachyose, and 0.0036−0.012 mg/g of verbascose. The average and median raffinose contents (0.079 and 0.078 mg/g) in almond milk were higher than stachyose (0.060 and 0.058 mg/g) and verbascose (0.0068 and 0.0055 mg/g). Soy milk consisted of 0.47−0.93 mg/g of raffinose, 2.9−5.4 mg/g of stachyose, and 0.18−0.31 mg/g of verbascose. Defatted soy flour contained 12−15 mg/g of raffinose, 56−59 mg/g of stachyose, and 3.3−3.7 mg/g of verbascose. Full-fat soy flour contained 5.6−10 mg/g of raffinose, 40−47 mg/g of stachyose, and 1.5−3.8 mg/g of verbascose (excluding the FFSF4 sample). FFSF4 was Korean fermented soybean powder ("mejugaru"), containing extraordinarily low raffinose (0.20 mg/g) and stachyose (3.6 mg/g) contents; oligosaccharides in soybeans were likely degraded or utilized by microorganisms during fermentation. Stachyose was the most abundant oligosaccharide in all soy milk and defatted and full-fat soy flour samples, followed by raffinose and verbascose. Raffinose and stachyose contents in different varieties of defatted soy flour determined in a previous study were 7.8–14.1 and 35.3–57.8 mg/g (dry basis), respectively (Bainy et al., 2008), which was in a similar range to our measured values (13–16 and 60–63 mg/g, on dry basis). In general, defatted soy flour contained more oligosaccharides than full-fat soy flour due to the lower lipid content in defatted soy flour (3.1−3.6% vs. 18.5−23.3%, as is) (FoodData Central, 2022); soy milk contained a lower amount of oligosaccharides than soy flour because of the higher moisture content (90.3−93.6% vs. 5.53−8.52%). The oligosaccharide contents in almond milk were lower than in soy milk, given the fact that almonds contained lower amounts of oligosaccharides (7.1–21.1 mg/g raffinose) than soybeans (8–13 mg/g raffinose, 32–43 mg/g stachyose, and 1–2 mg/g verbascose) (Barreira et al., 2010; Fan et al., 2015; Kuo et al., 1988).

Validation data for the batch analysis are presented in Table 1. All the coefficients of determination (r^2) of the quadratic calibration curves were above 0.99999. The low intra- (0.0−3.8%) and inter-batch RSD (1.3−5.2%) verified the precision of the measurements. The instrumental LOD were 1.2, 1.3, and 3.4 μg/L for raffinose, stachyose, and verbascose, respectively, which represent a considerable improvement over the values recently reported in the literature (49.63, 17.08, and 2891.22 μg/L, respectively) (Pico et al., 2021); the instrumental LOQ were 4.9, 6.3, and 12 μg/L, respectively.

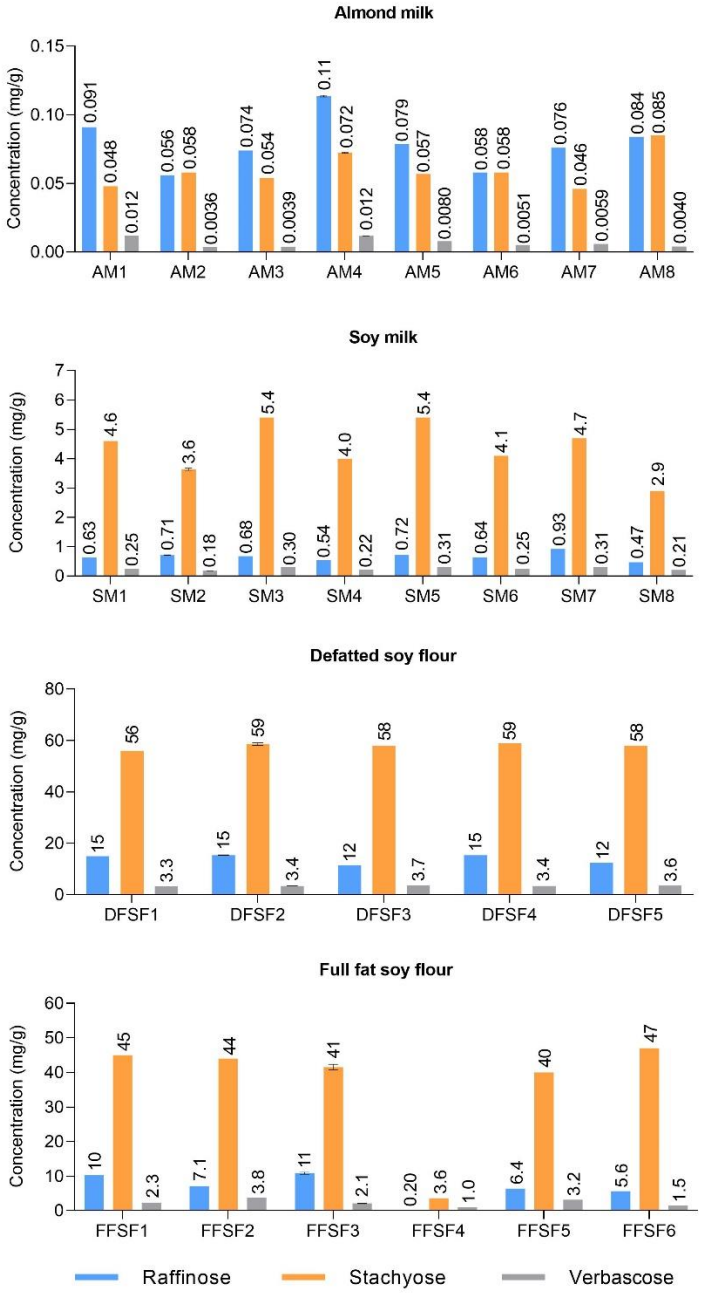


Fig. 5.2. Concentrations of raffinose, stachyose, and verbascose in commercial almond milk (AM1–AM8), soy milk (SM1–SM8), and soy flour (DFSF1–DFSF5 (defatted) and FFSF1–FFSF6 (full-fat)) products measured by HPAE-PAD. Data of AM4 (n = 3), SM2 (n = 4), DFSF2 (n = 2), and FFSF3 (n = 2) were expressed as mean ± standard deviation to show the measurement uncertainty; data of the other samples (n = 1) represent the measured values.

5.3.2. Comprehensive oligosaccharide profiling by LC-Q-TOF MS

LC-MS/MS analysis was carried out to identify the monosaccharide composition of the many oligosaccharides for which commercial standards are not yet available. To encompass the variety of oligosaccharides and keep the data analysis workload manageable, different brands of each sample type were pooled prior to LC-MS/MS analysis. The results revealed that almond milk, soy milk, and soy flour contained a variety of oligosaccharides besides the three major oligosaccharides (Table 5.S4). A total of 82, 60, 48, and 75 oligosaccharides were identified from the pooled almond milk, soy milk, defatted-, and full-fat soy flour samples, respectively. Most of the oligosaccharides identifications were confirmed in at least one pooled sample by inspecting the fragment ion peaks in the tandem MS spectra. A majority of the identified oligosaccharides (47, 40, 29, and 51, respectively) comprised only hexoses, with a degree of polymerization of 2–8. Five oligosaccharides containing pentose units (Hex_4Pent_1, Hex_5Pent_1, and Hex_3Pent_2) were identified (three confirmed by tandem MS confirmation, Supplementary material Fig. 5.S1A) in the pooled full-fat soy flour.

Some oligosaccharides contained residues other than common monosaccharides, such as pinitol, phosphoryl group, and acetyl group, according to the mass-to-charge ratios (*m/z*) observed in the LC-MS/MS analysis. Ciceritol, a pinitol digalactoside, was identified in all the four pooled samples (an example tandem MS spectrum was shown in Fig. 5.S1B). Ciceritol is present in chickpeas and lentils in high abundances (Quemener & Brillouet, 1983) and was found to exert *in vitro* prebiotic activity in a previous study (Zhang et al., 2017). Ciceritol was also found in soybean in previous studies, while its concentration (0.008 mg/g) was much lower than chickpea (0.280 mg/g) and lentil (0.160 mg/g) (Obendorf, Horbowicz, Dickerman, Brenac, & Smith, 1998; Quemener & Brillouet, 1983). In the current study, ciceritol found in the pooled almond milk

sample appeared to be even less abundant than in the pooled soy milk sample, considering the injection volumes and peak areas. To the best of our knowledge, ciceritol was identified in almond products for the first time.

Phosphorylated oligosaccharides were found in all four pooled samples. The three pooled soy samples all contained ten phosphorylated oligosaccharides (four of Hex_3P_1 (*m/z* 585.143, $[M + H]^+$) and six of Hex_4P_1 (*m/z* 747.196, $[M + H]^+$)). The pooled almond sample contained 12 phosphorylated oligosaccharides (six Hex_3P_1 and six Hex_4P_1), among which ten were determined to be the same as those found in the soy samples based on the retention times. For the phosphorylated oligosaccharides fragmented by CID, strong signal of the fragment ions, Hex_1P_1 (*m/z* 243.027, $[Hex_1P_1 - H_2O + H]^+$), Hex_2P_1 (*m/z* 405.080 $[Hex_2P_1 - H_2O + H]^+$), and Hex_3P_1 (*m/z* 567.132, $[Hex_3P_1 - H_2O + H]^+$, from Hex_4P_1 molecules) (all with a neutral loss of water) on the tandem MS spectra (Fig. 5.3A) supported their identifications. Acetylated oligosaccharides were found in all the soy and almond samples. Five acetylated oligosaccharides containing four hexoses and one acetyl group ($Hex_3HexOAc_1$, *m/z* 726.266, $[M + NH_4]^+$) were identified from soy milk, defatted soy flour, and full-fat soy flour with tandem MS confirmation (Fig. 5.S1C). Three of them were also found in common in the almond milk sample.

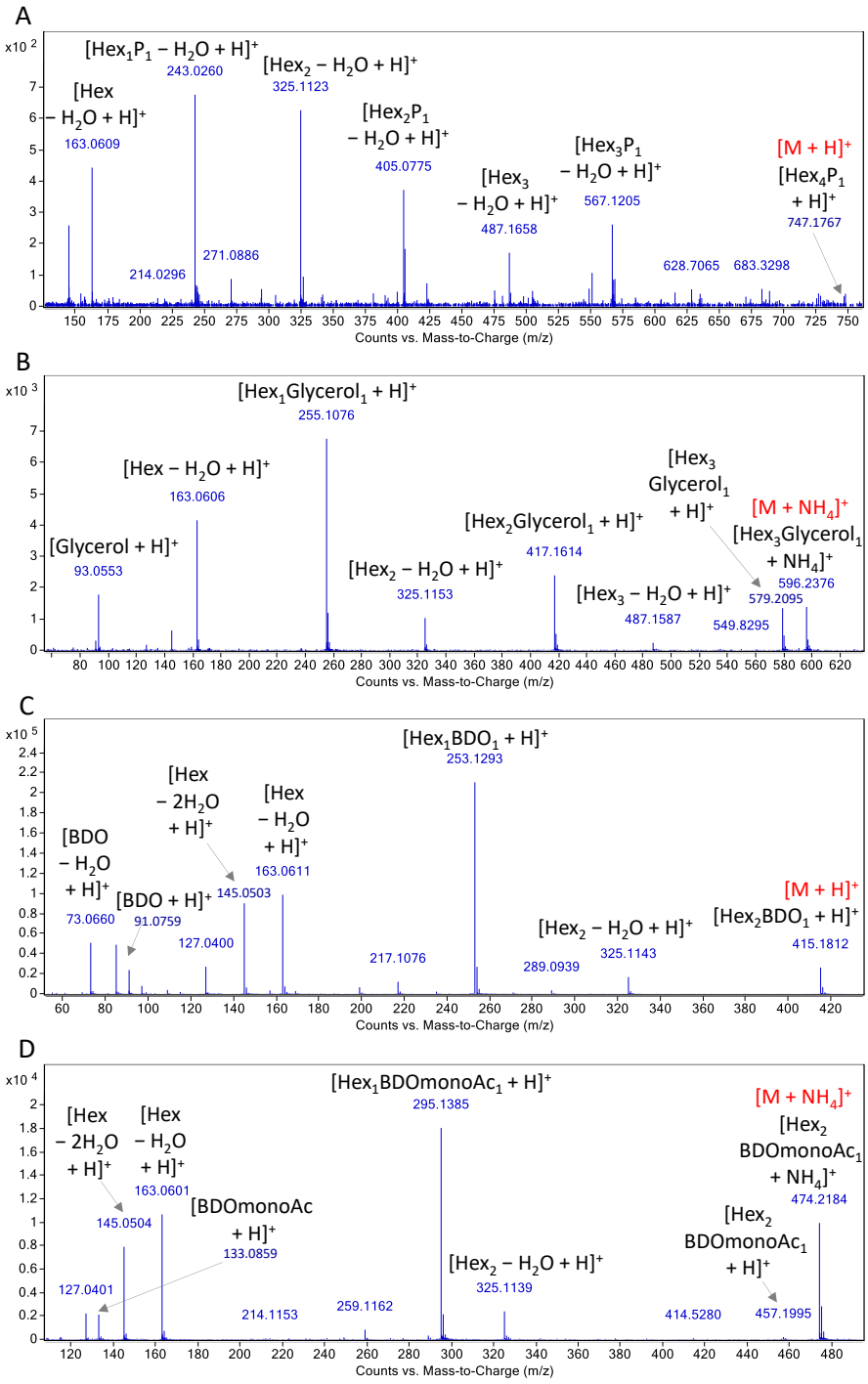


Fig. 5.3. LC-Q-TOF tandem MS spectra of Hex_4P_1 (13.349 min; spectrum of soy milk; A), $Hex_3Glycerol_1$ (10.002 min; spectrum of full-fat soy flour; B), and Hex_2BDO_1 (16.623 min; C) and $Hex_2BDOmonoAc_1$ l (18.918 min; D) in almond milk generated from precursor ions of $[M + H]^+$ or $[M + NH_4]^+$ (annotated in red). BDO: butanediol; BDOmonoAc: butanediol monoacetate.

A series of oligosaccharides with the *m/z* of 434.187, 596.240, and 758.292 were tentatively identified as glycerol-containing oligosaccharides. The fragment ions *m/z* 93.05 ($[glycerol + H]^+$), *m/z* 205.107 ($[Hex_1Glycerol_1 + H]^+$), *m/z* 417.160 ($[Hex_2Glycerol_1 + H]^+$), and *m/z* 579.213 ($[Hex_3Glycerol_1 + H]^+$) in the tandem MS spectra (Fig. 5.3B) of this series of oligosaccharides indicates that glycerol constitutes part of the oligosaccharides. The matched retention times of the oligosaccharides consisting of a glycerol and three or four hexose residues found in the almond and soy samples suggest that almond and soy can synthesize some identical glycerol-containing oligosaccharides. Glycerol-containing oligosaccharides were previously identified in wheat flour and algae (Carter, McCluer, & Slifer, 1956; Karsten, Michalik, Michalik, & West, 2005). Glyceryl glycosides with only one monosaccharide conjugated with glycerol were also found in algae, wine, and sake (Eggert & Karsten, 2010; Ruiz-Matute, Sanz, Moreno-Arribas, & Martínez-Castro, 2009). However, this type of compounds had not been previously reported in soy and almond products.

Noteworthy, several oligosaccharides containing 2–3 hexose units and an unknown residue were exclusively detected in the almond milk sample, e.g., three peaks whose protonated molecules ($[M + H]^+$) had an *m/z* of 415.183. The areas of the three peaks were much larger than verbascose in almond milk, suggesting the significant abundance of these compounds in almond milk. The unknown residues were suggested to be conjugated to the carbohydrate moieties by a glycosidic linkage (i.e., the oligosaccharides were glycosides with the unknown residues). According to the LC-MS/MS data, the monoisotopic masses of the aglycones of the two different series of glycosides were 90.068 and 132.079. The fragment ions of *m/z* 73.066 ($[M – Hex_2 – H_2O + H]^+$) and *m/z* 91.076 ($[M – Hex_2 + H]^+$) from the precursor ion of *m/z* 415.183 ($[M + H]^+$) (Fig. 5.3C) were associated with the unknown residue with a monoisotopic mass of 90.068. Another

nine peaks with *m/z* 577.234 ($[M + H]^+$) and *m/z* 594.261 ($[M + NH_4]^+$) also had the fragment ions of *m/z* 73.066 ($[M – Hex_2 – H_2O + H]^+$) and *m/z* 91.076 ($[M – Hex_2 + H]^+$), indicating that they possess the same unknown residue with the *m/z* 415.183 peaks. Similarly, the product ion series of *m/z* 133.086 ($[M – Hex_2 + H]^+$), *m/z* 295.139 ($[M – Hex + H]^+$), and *m/z* 457.200 ($[M + H]^+$), with a mass interval of ~162.05, from the precursor ion of *m/z* 474.220 ($[M + NH_4]^+$) (Fig. 5.3D) were related to the unknown residue of 132.079. The fragment ion peaks of *m/z* 325.114 corresponding to Hex_2 were found in the tandem MS spectra from both the precursors of *m/z* 415.183 and *m/z* 474.220, indicating that each of the unknown residues was linked to a disaccharide with two hexose units. Likewise, fragment ion peaks of *m/z* 487.168 ($[Hex_3 – H_2O + H]^+$) from the precursors of *m/z* 577.234 and *m/z* 594.261 demonstrated the presence of a trisaccharide with three hexose units in their structures.

Because the oligosaccharide samples were purified with mixed-mode and PGC SPE in series, which should remove most compounds with hydrophobic moieties, the unknown residues were expected to be highly polar. Based on the monoisotopic masses and the expected physicochemical properties, the two aglycone residues might be butanediol (90.068) and butanediol acetate (132.079). 2,3-Butanediol is a known sensory compound found in non-bitter almonds (Garg et al., 2018; Wirthensohn et al., 2008). 2,3-Butanediol acetate is a volatile compound found in wine (Wyk, Kepner, & Webb, 1967) and muskmelon (Lignou, Parker, Oruna-Concha, & Mottram, 2013), but it has not been identified in almond to date. Although in theory, other butanediol isomers (e.g., 1,4-butanediol and 1,3-butanediol) could also correspond to the unknown residue (90.068) or part of the aglycone residue (132.079), because of the existence of 2,3-butanediol in almonds, we tentatively identified the residues as 2,3-butanediol (90.068) and 2,3-butanediol acetate (132.079). While 2,3-butanediol is found in free form in almonds, it is

plausible that 2,3-butanediol could be incorporated into other compounds thanks to various metabolic pathways. For example, in previous studies, 2,3-butanediol glucoside was found in fennel (Kitajima, Ishikawa, & Tanaka, 1998) and *in vitro* fecal fermentation product of black rice (Owolabi, Dat-arun, Takahashi Yupanqui, & Wichienchot, 2020). 2,3-Butanediol was also found to be conjugated with a disaccharide glycoside to form 2,3-butanediol apiosyl-glucoside (Kitajima et al., 1998). The actual structures of the compounds tentatively identified as glycosides of 2,3-butanediol and 2,3-butanediol acetate in almonds would require further investigation, which is beyond the scope of this work.

Fig. 5.4 shows the apparent relative abundance of various oligosaccharides identified by LC-Q-TOF MS. Interestingly, minor oligosaccharides in almond milk accounted for 25% of the total, bringing them close in abundance to stachyose (29%), the second most represented oligosaccharide. Among the minor ones, the glycosides of 2,3-butanediol (Hex_{2-3} + butanediol) was the highest class (19%), followed by Hex_{3-8} (5%). Among the three pooled soy samples, oligosaccharides' distribution was similar. Minor oligosaccharides accounted for 6–10%, with $Hex_{3-6/7}$ as the most represented, followed by ciceritol.

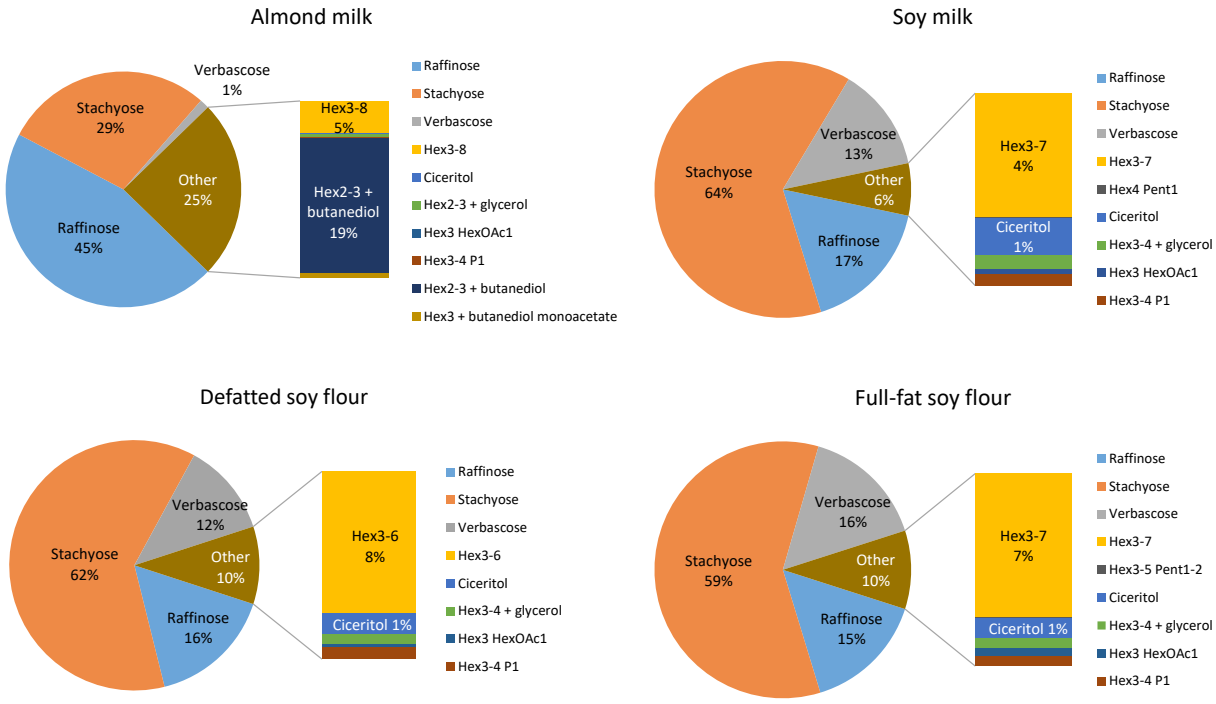


Fig. 5.4. Apparent relative abundance of different classes of oligosaccharides identified in the almond milk, soy milk, and soy flour estimated by peak areas from the LC-Q-TOF analysis. $Hex_{3\text{-}6/7/8}$ group excludes raffinose, stachyose, and verbascose. Hex: hexose; Pent: pentose; HexOAc: acetyl-hexose; P: phosphorylation.

5.3.3. Identity confirmation of 2,3-butanediol glycosides

5.3.3.1. Enzymatic treatment using glucosidases

The purified almond milk oligosaccharide sample was treated with α-glucosidase and β-glucosidase (β-D-glucosidase) to examine the linkage type between different residues in the protonated molecule of *m/z* 415.183. The LC-QTOF analysis results showed that all three peaks in the extracted ion chromatograms (EIC) of *m/z* 415.183 were significantly reduced in the sample after treatment with β-glucosidase compared with the control (Fig. 5.5A), whereas those same peaks were not affected by α-glucosidase (Fig. 5.S2A and B). Also, as expected, the peak areas of

the major oligosaccharides, including raffinose, stachyose, and verbascose (containing α-galactosyl and α-1,β-2-glycosidic linkages), were not altered after the enzymatic treatment (Fig. 5.5B). These results revealed that the terminal hexose residue, and possibly the hexose residue attached to the aglycone, in the three compounds (*m/z* 415.183) was a β-D-glucose.

We hypothesized that the compounds with a protonated form of *m/z* 577.234 and/or an ammonium ion of *m/z* 594.261 contained the same aglycone as the *m/z* 415.183 compounds due to the identical mass of the unknown residue. However, the areas of the *m/z* 577.234 and *m/z* 594.261 peaks of the β-glucosidase-treated and the control samples were similar, suggesting that the terminal hexose was not a β-D-glucose. Interestingly, the *m/z* 474.220 peak ($[M + NH_4]^+$) with the proposed structure of 2,3-butanediol acetate glycoside, which contains two hexose units, completely disappeared after β-glucosidase treatment (Fig. 5.S2C), clearly indicating that at least one terminal hexose unit in this compound was a β-D-glucose.

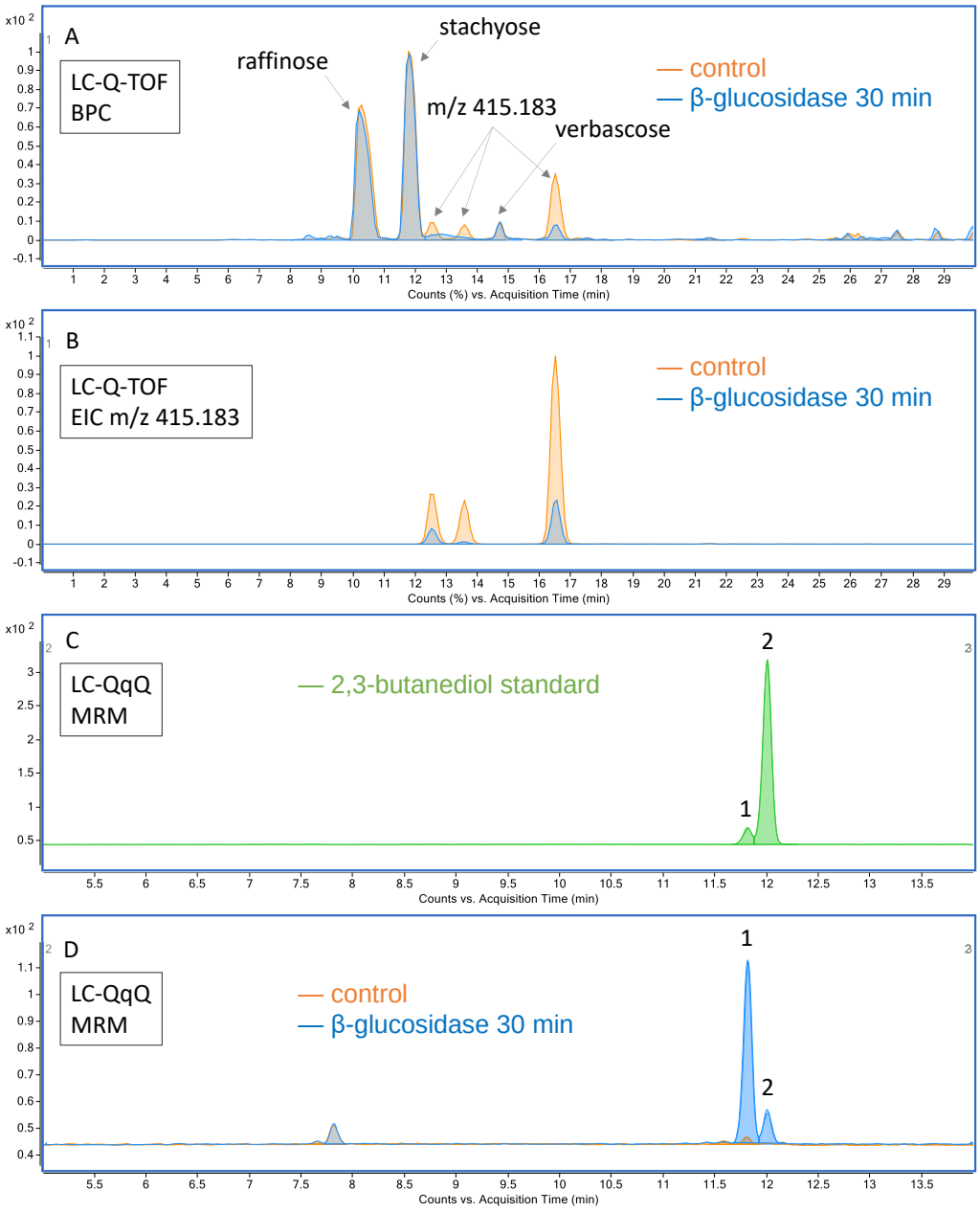


Fig. 5.5. Overlaid LC-Q-TOF chromatogram of purified almond neutral oligosaccharides undergoing β-glucosidase treatment (A, extracted ion chromatogram (EIC) of *m/z* 415.183; B, base peak chromatogram (BPC)) showing that the *m/z* 415.183 ion ($[M + H]^+$) peaks decreased after the treatment. Overlaid LC-QqQ MRM chromatograms (*m/z* 484 → 260 and *m/z* 484 → 262) of 2,3-butanediol standard (C) and almond neutral oligosaccharides with β-glucosidase treatment (D).

5.3.3.2. Analysis of the enzymatically released aglycone

2,3-Butanediol was derivatized using trichloroacetyl isocyanate to improve chromatographic retention and electrospray ionization in the LC-QqQ analysis. The β-glucosidase-treated almond oligosaccharide sample undergone the derivatization process was injected into the LC-QqQ to further confirm the identity of the enzymatically released aglycone.

The 2,3-butanediol standard had a minor peak at 11.82 min and a major peak at 12.01 min (Fig. 5.5C, peaks 1 and 2, respectively), both of which had nearly equivalent peak areas between the two MRM transitions (m/z 484 → 260 and m/z 484 → 262). According to the information provided by the manufacturer (MilliporeSigma), the 2,3-butanediol standard used in the current study is a mixture consisting 94.9% racemic and 4.8% meso forms. The percentages are similar to the area percentage of peaks 2 (92.1%) and 1 (7.9%), respectively. Therefore, peaks 1 and 2 might represent the meso and the racemic forms, respectively. Similarly, the β-glucosidase-treated sample contained two peaks (1 and 2), with similar peak areas between the two MRM transitions, at the same retention times as the standard. In comparison, the almond oligosaccharide sample without β-glucosidase treatment (control) only had a tiny peak at 11.81 min (Fig. 5.5D), confirming that β-glucosidase released the two peaks of 2,3-butanediol from the purified almond oligosaccharides.

In the β-glucosidase-treated sample, peak 1 (84.5%) was more abundant than peak 2 (15.5%), indicating that 2,3-butanediol released from the glycosides in almonds included more meso form than racemic form. (Wirthensohn et al., 2008) found that the ratio of racemic to mseo-2,3-butanediol in non-bitter almonds was 3.72 to 1. The abundance order of the two isomers of free 2,3-butanediol was contrary to the 2,3-butanediol released by β-glucosidase from purified

almond oligosaccharides. The cause of the inverse relative abundance of 2,3-butanediol isomers between the free and glycoside forms would need further investigation.

Since the release of 2,3-butanediol was corroborated with the LC-QqQ analysis by comparing with the authentic standard, it can be concluded that the almond milk sample contained at least one 2,3-butanediol-β-D-glucosyl-β-D-glucoside. The multiple *m/z* 415.183 peaks in the LC-Q-TOF chromatogram of the isomers might include different hexoses in the middle position, have the terminal β-glucose linked to different carbons on the innermost hexose residue, or contain different 2,3-butanediol stereoisomers. The structure of 2,3-butanediol-β-D-glucosido-β-D-glucoside is, to a certain degree, similar to amygdalin (D-mandelonitrile-β-D-glucosido-6-β-D-glucoside), which is more abundant in bitter almonds than sweet almonds. The biosynthesis and hydrolysis of amygdalin involve various specific enzymes (Thodberg et al., 2018). The naturally occurring β-glucosidase in almonds, which was also selected in the current study for the structure elucidation, is likely involved in the metabolism of the 2,3-butanediol glycosides in almonds. The findings of the 2,3-butanediol glycosides in almonds necessitate future studies on the role of these compounds in plant metabolism as well as in human nutrition.

5.4. Conclusions

Non-digestible oligosaccharides could beneficially affect human health due to their prebiotic activity. Therefore, it is essential to understand their composition and abundance in dietary sources. The current study optimized a quantification method for raffinose, stachyose, and verbascose using HPAE-PAD in almond milk, soy milk, and soy flour. The extraction of oligosaccharides was optimized by using water along with Carrez solutions to maximize the recovery and streamline the sample preparation procedures. The concentrations of raffinose, stachyose, and verbascose were surveyed in commercial almond milk, soy milk, and soy flour

obtained from different manufacturers. The additional analysis by LC-Q-TOF MS allowed the identification of many oligosaccharides with various structures, such as $Hex_{3–8}$, ciceritol, and $Hex_{2–3}Glycerol_1$, in these commercial products. Additionally, 2,3-butanediol glycosides containing one to two β-D-glucose residues were identified, for the first time, in almond milk in substantial relative abundance.

The quantification results presented here can serve to estimate oligosaccharide consumption from dietary intake. The data are novel and extend information about components in food reported in the USDA FoodData Central. Further investigation into the bioactivity of the newly identified oligosaccharides and glycosides is necessary to understand their role in human health.

CRediT author contribution statement

Yu-Ping Huang: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Visualization; Writing - original draft. **Bruna Paviani:** Investigation; Writing - review & editing. **Naomi K. Fukagawa:** Conceptualization; Funding acquisition; Project administration; Writing - review & editing. **Katherine Phillips:** Conceptualization, Methodology; Writing - review & editing. **Daniela Barile:** Conceptualization; Funding acquisition; Project administration; Supervision; Writing - review & editing.

Acknowledgments

The authors thank the USDA Beltsville Human Nutrition Research Center FoodData Central staff and Nancy Pennington and Ryan McGinty at Virginia Tech for assisting with study design and providing samples and Dr. Austin Horng-En Wang at the University of Nevada, Las Vegas for providing technical support in statistical programming with R. This work was supported by U.S.

Department of Agriculture Agricultural Research Service (USDA-ARS) (NACA #58-8040-8-013).

Supplementary material

Table 5.S1. Sample information and assay batch arrangement.

Sample No	Code	Sample description	Sample type	Assay Batch	Comments on assay	Ingredient list from product label
NFY121ZSM	SM1	Soy milk, unsweetened, plain, shelf stable	Soy milk	1		Water, organic soybeans
NFY121ZU5	SM2	Soy milk, unsweetened, plain, shelf stable	Soy milk	1*	*assay in duplicate	Organic soymilk (Filtered water, organic soybeans), Vitamin and mineral blend (calcium carbonate, vitamin A palmitate, vitamin D2, riboflavin [b2], vitamin B12), Sea salt, Natural flavor, Gellan gum.
NFY121ZU6	SM2	Soy milk, unsweetened, plain, shelf stable	Soy milk	2*	*assay in duplicate	Organic soymilk (Filtered water, organic soybeans), Vitamin and mineral blend (calcium carbonate, vitamin A palmitate, vitamin D2, riboflavin [b2], vitamin B12), Sea salt, Natural flavor, Gellan gum.
NFY121ZV0	SM3	Soy milk, unsweetened, plain, shelf stable	Soy milk	1		Purified water, organic soybeans
NFY121ZVZ	SM4	Soy milk, unsweetened, plain, shelf stable	Soy milk	2		Organic Soymilk (Filtered Water, Whole Organic Soybeans), Tricalcium Phosphate, Sea Salt, Organic Natural Flavor, Sodium Citrate, Carrageenan, Sodium Bicarbonate, Vitamin A Palmitate, Ergocalciferol (Vitamin D2)
NFY121ZXB	SM5	Soy milk, unsweetened, plain, shelf stable	Soy milk	2		Water, organic soybeans
NFY121ZYB	SM6	Soy milk, unsweetened, plain, shelf stable	Soy milk	1		Organic Soymilk (Filtered Water, Whole Organic Soybeans), Tricalcium Phosphate, Sea Salt, Natural Flavors, Sodium Citrate, Carrageenan, Sodium Bicarbonate, Vitamin A Palmitate, Vitamin D2 (Ergocalciferol)
NFY121ZZJ	SM7	Soy milk, unsweetened, plain, shelf stable	Soy milk	1		Filtered water, whole organic soybeans, tricalcium phosphate, sea salt, sodium citrate, carrageenan, organic natural vanilla flavor, natural flavor, vitamin A palmitate, vitamin D2
NFY12200Q	SM8	Soy milk, unsweetened, plain, shelf stable	Soy milk	2		Organic soymilk (filtered water, whole organic soybeans), calcium carbonate, organic locust bean gum, sea salt, natural flavors, gellan gum, vitamin A palmitate, vitamin D2, riboflavin (vitamin B2), vitamin B12.
NFY122027	AM1	Almond milk, unsweetened, plain, shelf stable	Almond milk	2		Almondmilk (Filtered Water, Almonds), Calcium Carbonate, Sea Salt, Potassium Citrate, Sunflower Lecithin, Gellan Gum, Natural Flavors, D-Alpha-Tocopherol (Natural Vitamin E).
NFY122039	AM2	Almond milk, unsweetened, plain, shelf stable	Almond milk	1		Water, ground almonds*, contains 1% or less of: gellan gum, natural flavor*, rice starch*, sea salt, vanilla extract*, vitamin extract*, vitamin D2, xanthan gum.
NFY12204B	AM3	Almond milk, unsweetened, plain, shelf stable	Almond milk	1		Almondmilk (filtered water, almonds), vitamin and mineral blend (calcium carbonate, vitamin E acetate, vitamin A palmitiate, vitamin D2), sea salt, gellan gum, locust bean gum, sodium ascorbate, natural flavor.

NFY12205H	AM4	Almond milk, unsweetened, plain, shelf stable	Almond milk	2		Almondmilk (Water, Almonds), Calcium Carbonate, Sunflower Lecithin, Sea Salt, Natural Flavors, Locust Bean Gum, Gellan Gum, Potassium Citrate.
NFY12205I	AM4	Almond milk, unsweetened, plain, shelf stable	Almond milk	2*	*assay in duplicate	Almondmilk (Water, Almonds), Calcium Carbonate, Sunflower Lecithin, Sea Salt, Natural Flavors, Locust Bean Gum, Gellan Gum, Potassium Citrate.
NFY12206B	AM5	Almond milk, unsweetened, plain, shelf stable	Almond milk	1		Almond Base, (Filtered Water, Almonds), Calcium Carbonate, Sea Salt, Sunflower Lecithin, Locust Bean Gum, Gellan Gum, Vitamin A Palmitate, Ergocalciferol (Vitamin D2), DL-Alpha-Tocopherol (Vitamin E), Riboflavin (Vitamin B2), Zinc Gluconate, Cyanocobalamin (Vitamin B12)
NFY12207I	AM6	Almond milk, unsweetened, plain, shelf stable	Almond milk	1		Almondmilk (Filtered Water, Almonds), Contains 2% or less of: Vitamin and Mineral Blend (Calcium Carbonate, Vitamin E Acetate, Vitamin A Palmitate, Vitamin D2), Sea Salt, Natural Flavor, Sunflower Lecithin, Locust Bean Gum, Gellan Gum, Ascorbic Acid
NFY12208P	AM7	Almond milk, unsweetened, plain, shelf stable	Almond milk	2		Almondmilk (Filtered Water, Almonds), Tricalcium Phosphate, Sea Salt, Sunflower Lecithin, Gellan Gum, Locust Bean Gum, Vitamin A Palmitate, Ergocalciferol (Vitamin D2), DL-Alpha Tocopherol Acetate (Vitamin E)
NFY12209U	AM8	Almond milk, unsweetened, plain, shelf stable	Almond milk	2		Organic Almondmilk (Filtered water, organic almonds), Sea Salt, Tricalcium Phosphate, Gellan Gum, Organic Locust Bean Gum, Vitamin A Palmitate, Ergocalciferol (Vitamin D2), DL-Alpha-Tocopherol Acetate (Vitamin E), Cyanocobalamin (Vitamin B12)
NFY121XV3	DFSF1	Flour, Soy (Defatted)	Soy flour, defatted	3		defatted soy flour
NFY121XVR	DFSF2	Flour, Soy (Defatted)	Soy flour, defatted	3*	*assay in duplicate	defatted soy flour
NFY121XWI	DFSF3	Flour, Soy (Defatted)	Soy flour, defatted	3		
NFY121XX4	DFSF4	Flour, Soy (Defatted)	Soy flour, defatted	3		defatted soy flour
NFY121XXT	DFSF5	Flour, Soy (Defatted)	Soy flour, defatted	3		
NFY121Y6F	FFSF1	Flour, Soy, Full Fat	Soy flour, full fat	3		organic soy
NFY121Y70	FFSF2	Flour, Soy, Full Fat	Soy flour, full fat	3		whole soya bean
NFY121Y7S	FFSF3	Flour, Soy, Full Fat	Soy flour, full fat	3*	*assay in duplicate	soya bean
NFY121Y8D	FFSF4	Flour, Soy, Full Fat (fermented soybean powder)	Soy flour, full fat	3		soy bean powder

NFY121Y95	FFSF5	Flour, Soy, Full Fat	Soy flour, full fat	3	100% organic soy flour
NFY121Y9T	FFSF6	Flour, Soy, Full Fat	Soy flour, full fat	3	soybeans
NFY121YA1	QC1	Soy flour CC	Soy flour, full fat	1	soybean flour
NFY1220AT	QC2	NIST3234 Soy Flour	Soy flour	1	
NFY121YA2	QC1	Soy flour CC	Soy flour, full fat	2	soybean flour
NFY1220AV	QC2	NIST3234 Soy Flour	Soy flour	2	
NFY121YA3	QC1	Soy flour CC	Soy flour, full fat	3	soybean flour
NFY1220AW	QC2	NIST3234 Soy Flour	Soy flour	3	

Table 5.S2. Comparison of the efficiency of oligosaccharide extraction from almond milk and soy milk between ethanol precipitation and Carrez clarification (n = 3).

	Raffinose	Stachyose	Verbascose
		mg/g	
Almond milk			
Ethanol precipitation (4V)[1]	0.085 ± 0.002[a2]	0.041 ± 0.001[b]	0.0036 ± 0.0001[a]
Ethanol precipitation (2V)	0.088 ± 0.003[a]	0.044 ± 0.001[a]	0.0037 ± 0.0002[a]
Carrez clarification	0.088 ± 0.001[a]	0.044 ± 0.000[a]	0.0037 ± 0.0000[a]
Soy milk			
Ethanol precipitation (4V)	0.78 ± 0.01[b]	3.8 ± 0.2[b]	0.25 ± 0.01[b]
Ethanol precipitation (2V)	0.85 ± 0.02[ab]	4.8 ± 0.1[a]	0.29 ± 0.01[a]
Carrez clarification	0.91 ± 0.06[a]	5.2 ± 0.3[a]	0.31 ± 0.01[a]

[1] 4V and 2V represent adding 4 and 2 volumes of ethanol, respectively.

[2] Different superscript letters indicate statistically significant differences across extraction methods by one-way ANOVA followed by Tukey's test at $p < 0.05$.

Table 5.S3. Comparison of the efficiency of oligosaccharide extraction from defatted soy flour using aqueous extraction and Carrez clarification (n = 3).

Extraction and clarification procedure[1]	Carrez I + Carrez II	Raffinose	Stachyose	Verbascose
	μL		mg/g	
Combined	100	9.1 ± 0.2	56 ± 1	3.3 ± 0.1
Separate	100	9.4 ± 0.4	58 ± 3	3.5 ± 0.2
Separate	200	9.2 ± 0.2	56 ± 1	3.5 ± 0.1

[1] Combined refers to mixing soy flour, water, and Carrez solutions together; separate refers to extracting soy flour with water and then clarifying the supernatant with Carrez solutions. There was no significant difference in the measured oligosaccharide contents when using different extraction and clarification procedures based on one-way ANOVA at $p < 0.05$.

Table 5.S4. Identification of oligosaccharides in almond milk, soy milk, and soy flour (pooled samples).

RT (min)	*m/z*				(Suggested) compound	Major fragments (*m/z*)	Almond milk		Soy milk		Soy flour, defatted		Soy flour, full-fat	
	$[M+H]^+$	$[M+NH4]^+$	$[M+Na]^+$	$[M+K]^+$			MS	MSMS	MS	MSMS	MS	MSMS	MS	MSMS
6.541	505.176	**522.204**	527.159	543.136	**Hex_3 (manninotriose)**	127.04, 145.05, 163.06, 289.09, 307.10, 325.11, 343.12, 487.17, 505.18					✓	✓	✓	✓
6.654	505.178	**522.204**	527.159	543.134	**Hex_3**	127.04, 145.05, 163.06, 289.09, 307.10, 325.11, 343.12, 487.17, 505.18	✓	✓						
6.704	505.176	**522.204**	527.159		**Hex_3**	127.04, 145.05, 163.06, 289.09, 307.10, 325.11, 343.12, 487.17, 505.18			✓	✓				
6.905	505.177	**522.204**	527.159	543.134	**Hex_3**	127.04, 145.05, 163.06, 289.09, 307.10, 325.11, 343.12, 487.17, 505.17							✓	✓
7.158	505.177	**522.205**	527.157	543.132	**Hex_3**	127.04, 145.05, 163.06, 289.09, 307.10, 325.11, 343.12, 487.17, 505.17	✓	✓						
7.309	505.177	**522.204**	527.160	543.134	**Hex_3 (manninotriose)**	127.04, 145.05, 163.06, 289.09, 307.11, 325.11, 343.12, 487.17, 505.18			✓	✓	✓	✓	✓	✓
7.665	505.180	**522.203**	527.158	543.140	**Hex_3**	127.04, 145.05, 163.06, 289.09, 307.10, 325.11, 343.13, 487.18, 505.18	✓	✓						

8.569	505.182	**522.204**	527.159	543.137	**Hex$_3$**	127.04, 145.05, 163.06, 289.09, 307.10, 325.11, 343.12, 487.16, 505.17	✓	✓						
8.893	505.180	**522.200**	527.161	543.137	**Hex$_3$**	145.05, 163.06, 181.07, 289.09, 307.10, 325.11, 343.12, 505.18			✓	✓	✓	✓	✓	✓
9.174	505.179	**522.204**	527.160	543.132	**Hex$_3$**	127.04, 145.05, 163.06, 289.09, 307.11, 325.11, 343.12, 487.17, 505.18	✓	✓						
9.880	505.181	522.204	527.160		Hex$_3$		✓							
10.543	505.178	**522.204**	527.161		**Hex$_3$ (raffinose)**	127.04, 145.05, 163.06, 181.07, 289.09, 307.10, 325.11, 343.12, 487.17, 505.18	✓	✓	✓	✓	✓	✓	✓	✓
11.837	505.178	**522.204**			**Hex$_3$**	127.04, 145.05, 163.06, 289.09, 307.10, 325.11, 343.12, 487.17, 505.18	✓	✓						
12.484	505.178	**522.204**			**Hex$_3$**	127.04, 145.05, 163.06, 289.09, 307.10, 325.11, 343.12, 487.17, 505.18	✓	✓						
12.522	505.178	**522.204**	527.163		**Hex$_3$**	145.05, 163.06, 181.06, 307.10, 325.11, 343.12, 487.17, 505.18			✓	✓	✓	✓	✓	✓
13.550	505.175	**522.203**	527.158	543.138	**Hex$_3$**	127.04, 145.05, 163.06, 325.12, 487.17, 505.16			✓	✓				
15.062	505.177	**522.206**			**Hex$_3$**	127.04, 145.05, 163.06, 325.11, 487.17, 505.16			✓	✓				
16.515	505.177	522.204			Hex$_3$		✓							

17.565	505.178	**522.204**			**Hex$_3$**	127.04, 145.05, 163.06, 325.11, 343.12, 487.17, 505.17			✓	✓			✓	
19.119	505.180	**522.204**			**Hex$_3$**	145.05, 163.06, 325.12, 343.13, 487.14, 505.19	✓	✓	✓	✓			✓	✓
19.623	505.177	**522.204**			**Hex$_3$**	145.05, 163.06, 325.11, 487.17, 505.18	✓	✓	✓	✓			✓	✓
8.771		**684.255**	689.212	705.185	**Hex$_4$**	145.05, 163.06, 325.11, 343.12, 487.16, 505.17, 667.24	✓	✓						
9.174		**684.250**	689.210	705.187	**Hex$_4$**	127.04, 145.05, 163.06, 289.09, 307.10, 325.11, 343.12, 487.17, 505.17, 649.22, 667.24	✓	✓						
9.577		684.257	689.210	705.185	Hex$_4$		✓							
10.888	667.231	684.253			Hex$_4$		✓							
11.291	667.229	**684.256**			**Hex$_4$**	127.04, 145.05, 163.06, 325.11, 343.12, 487.17, 505.17, 649.22, 667.23	✓	✓						
8.830	667.225	**684.257**	689.214	705.174	**Hex$_4$**	145.05, 163.06, 325.11, 343.11, 487.16, 649.22, 667.23			✓	✓				
9.097	667.217	**684.254**	689.212	705.181	**Hex$_4$**	163.06, 325.11, 343.13, 487.16, 505.17					✓	✓	✓	✓
9.938	667.234	684.254	689.207	705.162	Hex$_4$				✓					
10.407	667.229	**684.257**	689.212	705.182	**Hex$_4$**	127.04, 145.05, 163.06, 325.11, 343.13, 487.16, 505.18, 649.22, 667.21					✓	✓	✓	

11.232	667.226	**684.256**	689.210	**Hex$_4$**	127.04, 145.05, 163.06, 325.11, 343.12, 487.17, 505.18, 649.22, 667.23					✓	✓	✓	✓
12.018	667.231	**684.260**		**Hex$_4$ (stachyose)**	145.05, 163.06, 325.12, 343.12, 487.17, 505.18, 649.22, 667.22	✓	✓	✓	✓	✓	✓	✓	✓
13.189	667.233	684.256		Hex$_4$		✓		✓	✓	✓	✓	✓	
14.096	667.230	**684.256**		**Hex$_4$**	145.05, 163.06, 325.11, 343.13, 487.17, 505.18, 649.22	✓	✓	✓	✓	✓	✓	✓	✓
14.701		**684.258**		**Hex$_4$**	145.05, 163.06, 325.11, 343.12, 487.17, 505.17, 649.22, 667.23	✓	✓	✓	✓	✓	✓	✓	✓
15.608	667.234	**684.257**		**Hex$_4$**	145.05, 163.06, 325.11, 343.12, 487.17, 505.17	✓	✓	✓	✓	✓	✓	✓	✓
16.213		684.256		Hex$_4$		✓							
16.738	667.234	**684.258**		**Hex$_4$**	145.05, 163.06, 325.12, 343.12, 487.17, 505.17							✓	✓
17.666	667.231	**684.257**		**Hex$_4$**	145.05, 163.06, 325.11, 343.12, 487.17, 505.18, 667.23	✓	✓	✓	✓	✓	✓	✓	✓
19.019		**684.255**		**Hex$_4$**	145.05, 163.06, 325.11, 343.13, 487.17, 505.18	✓	✓						
21.337	667.234	**684.257**		**Hex$_4$**	145.05, 163.06, 325.11, 343.12, 487.17, 505.18, 667.24	✓	✓						
25.110	667.230	**684.251**	689.199	**Hex$_4$**	163.06, 325.11, 343.13, 487.16, 505.18, 649.22, 667.23			✓	✓				

26.354	667.223	**684.256**	689.212	**Hex$_4$**	163.06, 325.11, 343.13, 487.16, 505.18, 649.22, 667.23			✓	✓	✓	✓			
27.059		684.252		Hex$_4$		✓		✓		✓				
27.462		**684.257**		**Hex$_4$**	163.06, 325.12, 505.17	✓	✓	✓	✓	✓	✓			
10.506		846.308		Hex$_5$						✓		✓		
11.211	829.281	**846.308**	851.263	**Hex$_5$**	163.06, 325.11, 343.13, 487.17, 505.18, 649.21, 667.22	✓	✓	✓	✓	✓	✓	✓	✓	
11.715		**846.309**		**Hex$_5$**	163.06, 325.11, 343.12, 487.16, 505.17, 649.22, 811.25							✓	✓	
12.421	829.288	**846.310**		**Hex$_5$**	163.06, 325.11, 487.17, 649.22, 811.27			✓	✓			✓	✓	
13.614	829.280	**846.308**	851.266	**Hex$_5$**	163.06, 325.11, 487.17, 505.18, 649.22, 811.27	✓	✓	✓	✓	✓	✓	✓	✓	
14.924	829.281	**846.311**		**Hex$_5$ (verbascose)**	163.06, 325.11, 487.17, 505.18, 649.22, 667.23, 811.27	✓	✓	✓	✓	✓	✓	✓	✓	
15.529	829.280	**846.309**	851.265	**Hex$_5$**	163.06, 325.11, 487.16, 667.23, 829.27	✓	✓	✓	✓	✓	✓	✓	✓	
16.134	829.288	**846.309**	851.261	**Hex$_5$**	163.06, 325.11, 487.17, 505.18, 649.22, 667.23	✓	✓	✓	✓	✓	✓	✓	✓	
16.738	829.285	**846.307**	851.269	**Hex$_5$**	163.06, 325.11, 487.16, 505.18, 649.21			✓	✓	✓	✓	✓	✓	
17.343	829.284	**846.310**		**Hex$_5$**	163.06, 325.11, 487.17, 505.17, 649.22	✓		✓		✓	✓	✓	✓	
18.149		**846.309**		**Hex$_5$**	163.06, 325.11, 487.17, 505.18, 667.23	✓				✓	✓	✓		

18.738		846.311		Hex_5		✓			✓		✓	
20.552		846.308		Hex_5							✓	
26.757	829.282	**846.308**		**Hex_5**	163.06, 325.11, 487.17, 649.22, 829.28		✓	✓				
27.647	829.277	**846.308**		**Hex_5**	163.06, 325.11, 487.16, 649.21, 829.28		✓	✓				
32.105	829.271	**846.305**		**Hex_5**	163.06, 325.11, 487.17, 505.18, 649.22, 667.23	✓	✓					
11.577		1008.359		Hex_6		✓			✓			
12.925		**1008.362**	1013.313	**Hex_6**	163.06, 325.11, 487.17, 649.21, 811.26						✓	✓
13.311		**1008.360**	1013.332	**Hex_6**	163.06, 325.11, 487.17, 649.22, 811.27						✓	✓
13.717		**1008.361**	1013.310	**Hex_6**	163.06, 325.11, 487.16, 649.24, 667.23				✓	✓		
13.815		**1008.361**		**Hex_6**	163.06, 325.11, 487.17, 649.24, 811.27						✓	✓
14.096		1008.360		Hex_6		✓						
14.256	991.347	1008.361		Hex_6			✓					
15.163	991.331	**1008.361**		**Hex_6**	163.06, 325.11, 487.17, 649.21, 811.27, 991.35		✓	✓				
15.327		**1008.359**		**Hex_6**	163.06, 325.11, 487.17, 649.21, 811.26						✓	✓
15.731		**1008.361**		**Hex_6**	163.06, 325.11, 487.16, 649.21, 811.28						✓	✓
16.070	991.329	**1008.362**	1013.321	**Hex_6**	163.06, 325.11, 487.17, 649.22, 811.27, 973.32		✓	✓				

16.421		**1008.357**	1013.323	**Hex$_6$**	163.06, 325.11, 487.16, 649.22, 829.27					✓	✓		
16.759	991.343	**1008.360**	1013.314	**Hex$_6$**	163.06, 325.11, 487.17, 649.22, 811.27			✓	✓			✓	✓
17.868		1008.359		Hex$_6$				✓					
17.242		**1008.362**	1013.319	**Hex$_6$**	163.06, 325.11, 487.17, 649.22, 667.23, 829.29					✓	✓	✓	✓
18.049	991.339	**1008.362**	1013.316	**Hex$_6$**	163.06, 325.11, 487.17, 505.18, 649.22, 667.23, 829.28							✓	✓
18.452		**1008.360**		**Hex$_6$**	163.06, 325.11, 487.16, 649.22, 667.23, 829.29					✓	✓	✓	✓
19.042		1008.360	1013.323	Hex$_6$						✓			
19.846		**1008.360**		**Hex$_6$**	163.06, 325.11, 487.17, 649.23, 667.23, 829.26							✓	✓
15.163	991.336	**1008.362**		**Hex$_6$**	163.06, 325.11, 487.17, 649.22, 811.27, 991.33	✓	✓						
16.818	991.338	1008.361		Hex$_6$		✓							
17.624	991.346	1008.365		Hex$_6$		✓							
11.392	1153.346	1170.415		Hex$_7$		✓							
16.314	1153.388	1170.413		Hex$_7$		✓							
16.919	1153.392	1170.412		Hex$_7$		✓							
17.631	1153.385	1170.412		Hex$_7$		✓							
11.392		1170.416		Hex$_7$								✓	
15.831		**1170.414**	1175.362	**Hex$_7$**	163.06, 325.11, 487.17, 505.17, 667.23, 829.29							✓	✓
17.948		1170.414		Hex$_7$								✓	
19.040	1153.384	**1170.416**	1175.373	**Hex$_7$**	163.06, 325.11, 487.16, 505.17,							✓	✓

						649.22, 811.27, 973.33, 1153.40								
19.645		**1170.412**	1175.378		**Hex$_7$**	163.06, 325.11, 487.17, 505.18, 649.22, 829.28, 991.32							✓	✓
16.356	1153.386	**1170.415**	1175.359		**Hex$_7$**	163.06, 325.12, 487.16, 649.22, 811.27, 973.32, 1135.36, 1153.37			✓	✓				
18.976		**1170.420**	1175.385		**Hex$_7$**	163.06, 325.11, 487.16, 649.20, 829.27, 991.34, 1135.39			✓	✓				
34.855	1153.387	**1170.411**	1175.381		**Hex$_7$**	163.06, 325.12, 487.17, 649.23, 829.27, 991.34, 1135.39			✓	✓				
17.826	1315.436	**1332.461**			**Hex$_8$**	163.06, 325.11, 487.17, 649.23, 811.27, 973.33, 1135.36, 1297.40	✓	✓						
9.700	**519.192**	536.220	541.173	557.145	**Hex$_2$Methyl-inositol$_1$ (ciceritol)**	163.06, 195.09, 325.115, 357.139, 519.191	✓	✓	✓	✓	✓	✓	✓	✓
5.865	417.161	**434.188**	439.144	455.122	**Hex$_2$+glycerol**	93.054, 163.060, 255.107, 325.117, 417.157	✓	✓						
10.182	417.162	434.190	439.144		Hex$_2$+glycerol		✓							
10.002	579.215	**596.241**	601.193	617.168	**Hex$_3$+glycerol**	93.055, 163.061, 255.108, 325.115, 417.161, 487.159, 579.209	✓	✓	✓	✓	✓	✓	✓	✓
12.925	741.262	**758.293**	763.247		**Hex$_4$+glycerol**	93.057, 163.058, 255.107, 325.112, 417.162, 487.169, 579.217, 649.227, 741.259	✓		✓	✓	✓	✓	✓	✓

13.614	**758.295**	763.243	**Hex_4+glycerol**	93.053, 163.058, 255.107, 325.113, 417.162, 487.165, 579.212, 649.196, 741.250	✓		✓	✓	✓	✓	✓	✓
12.824	**726.267**	731.220	**$Hex_3HexOAc_1$**	163.06, 205.07, 325.11, 367.13, 487.17, 505.18, 529.18	✓		✓	✓	✓	✓	✓	✓
13.916	**726.266**		**$Hex_3HexOAc_1$**	163.06, 205.07, 325.12, 367.12, 529.17			✓		✓		✓	✓
14.823	**726.265**		**$Hex_3HexOAc_1$**	163.06, 205.07, 325.11, 367.12, 487.16, 547.19	✓		✓		✓	✓	✓	
16.638	**726.266**		**$Hex_3HexOAc_1$**	163.06, 205.07, 325.12, 349.11, 367.12, 529.19, 547.19			✓	✓	✓		✓	✓
17.746	**726.266**		**$Hex_3HexOAc_1$**	163.06, 205.07, 325.11, 349.11, 367.13, 487.17, 511.16, 529.17, 547.19	✓		✓		✓	✓	✓	✓
11.615	**816.300**	821.260	**Hex_4Pent_1**	133.29, 163.06, 295.10, 325.11, 343.117, 457.15, 487.17, 505.18, 619.21, 781.28							✓	✓
12.522	816.290		Hex_4Pent_1								✓	
15.327	816.290		Hex_4Pent_1				✓				✓	
13.513	**978.351**		**Hex_5Pent_1**	145.04, 163.06, 295.10, 325.11, 343.13, 457.15, 487.15, 505.18, 619.21, 781.26, 799.28, 961.32							✓	✓
25.845	**786.287**	791.233	**Hex_3Pent_2**	181.09, 295.10, 325.11, 457.16, 505.16, 589.19, 619.21, 751.23							✓	✓

11.291	585.143			Hex_3P_1		✓							
13.996	585.143			Hex_3P_1		✓							
14.622	**585.143**			**Hex_3P_1**	243.028, 405.077, 567.122	✓		✓	✓	✓	✓	✓	✓
15.025	585.143			Hex_3P_1		✓		✓		✓		✓	
15.932	585.142			Hex_3P_1		✓		✓		✓		✓	
17.041	**585.143**			**Hex_3P_1**	243.024, 405.068, 567.129	✓		✓		✓	✓	✓	✓
13.349	**747.196**			**Hex_4P_1**	145.045, 163.061, 243.026, 325.112, 405.077, 487.166, 567.121	✓		✓	✓	✓		✓	✓
13.752	**747.197**	764.224	769.177	**Hex_4P_1**	163.060, 243.026, 325.113, 405.081, 505.177, 567.131	✓		✓	✓	✓	✓	✓	✓
14.457	**747.194**	764.221	769.175	**Hex_4P_1**	243.028, 325.105, 405.077, 567.138	✓		✓	✓	✓	✓	✓	✓
15.768	**747.193**	764.223	769.176	**Hex_4P_1**	145.059, 243.024, 423.089, 567.137, 585.150	✓		✓	✓	✓	✓	✓	
17.041	**747.200**	764.221		**Hex_4P_1**	145.042, 163.064, 243.030, 405.076, 423.101, 567.130, 585.151	✓		✓		✓	✓	✓	
17.767	**747.196**			**Hex_4P_1**	243.024, 261.039, 423.083, 585.142	✓		✓		✓	✓	✓	
12.685	**415.182**	432.208		**Hex_2+2,3-butanediol**	73.066, 91.076, 163.06, 239.114, 253.129, 325.11, 401.165, 415.181	✓	✓						
13.693	**415.183**	432.209		**Hex_2+2,3-butanediol**	73.066, 91.076, 163.06, 253.129, 325.11	✓	✓						
16.623	**415.184**	432.210		**Hex_2+2,3-butanediol**	73.066, 91.076, 163.06, 253.130, 325.11, 415.181	✓	✓						
18.128	577.234	**594.261**		**Hex_3+2,3-butanediol**	73.063, 145.050, 163.060, 253.130, 325.113, 415.182, 487.164, 577.232	✓	✓						

19.019	577.232	594.261		Hex_3+2,3-butanediol		✓	
19.523	577.234	**594.259**	599.214	**Hex_3+2,3-butanediol**	145.046, 163.061, 253.129, 325.116, 415.178, 487.167, 577.223	✓	✓
19.926	577.232	**594.257**	599.217	**Hex_3+2,3-butanediol**	91.074, 145.052, 163.061, 253.130, 325.116, 415.179, 487.168, 577.235	✓	✓
21.639	577.235	**594.261**	599.216	**Hex_3+2,3-butanediol**	73.067, 91.076, 145.051, 163.061, 253.128, 325.114, 415.183, 487.168, 577.231	✓	✓
22.446	577.235	**594.261**	599.216	**Hex_3+2,3-butanediol**	73.067, 91.076, 145.051, 163.061, 253.128, 325.114, 415.183, 487.168, 577.231	✓	✓
24.361	577.237	594.264		Hex_3+2,3-butanediol		✓	
26.175	**577.234**	594.262	599.217	**Hex_3+2,3-butanediol**	73.065, 91.075, 145.049, 163.061, 253.128, 325.116, 415.184, 487.168, 577.231	✓	✓
26.780	**577.234**	594.262	599.217	**Hex_3+2,3-butanediol**	73.066, 91.077, 145.051, 163.061, 253.128, 325.113, 415.167, 487.160, 577.231	✓	✓
18.918		**474.219**	479.174	**Hex_2+2,3-butanediol monoacetate**	133.086, 145.05, 163.06, 295.138, 325.11, 457.20	✓	✓
24.663		**636.271**	641.228	**Hex_3+2,3-butanediol monoacetate**	133.090, 145.052, 163.061, 295.141, 325.114, 457.192, 487.162	✓	✓

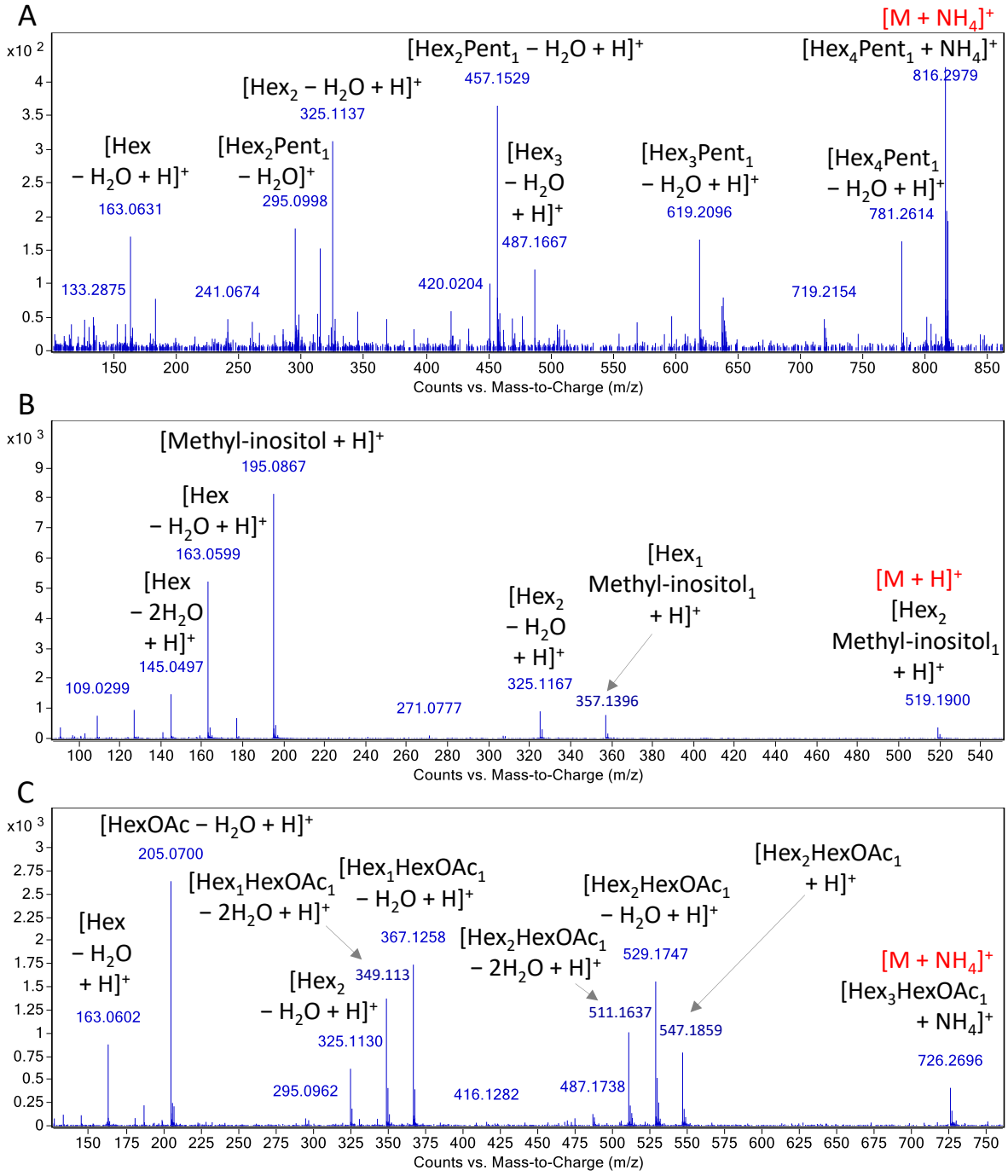


Fig. 5.S1. LC-Q-TOF tandem MS spectra of Hex_4Pent_1 (11.615 min; spectrum of full-fat soy flour; A) ciceritol (9.700 min; spectrum of soy milk; B), and $Hex_3HexOAc_1$ (17.746 min; spectrum of full-fat soy flour; C), generated from precursor ions of $[M + H]^+$ or $[M + NH_4]^+$ (annotated in red).

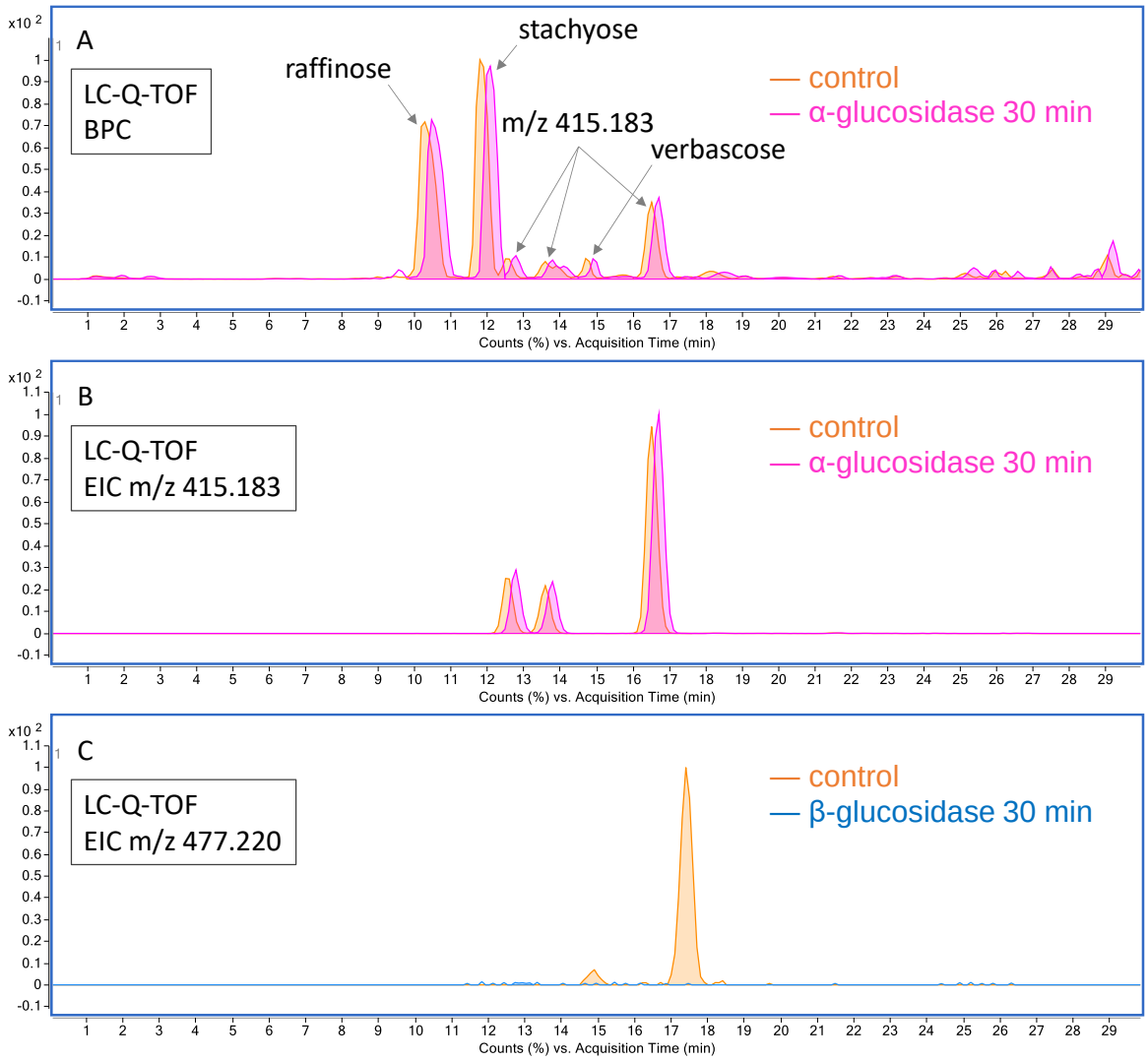


Fig. 5.S2. Overlaid LC-Q-TOF chromatogram of purified almond neutral oligosaccharides undergoing α-glucosidase treatment (A, base peak chromatogram (BPC); B, extracted ion chromatogram (EIC) of *m/z* 415.183) showing that the *m/z* 415.183 ion ($[M + H]^+$) peaks did not decrease after the treatment. Overlaid LC-Q-TOF chromatogram of purified almond neutral oligosaccharides undergoing β-glucosidase treatment (C, EIC of *m/z* 477.220) showing that the *m/z* 477.220 ion ($[M + NH_4]^+$) peak disappeared after the treatment.

References

Aldredge, D. L., Geronimo, M. R., Hua, S., Nwosu, C. C., Lebrilla, C. B., & Barile, D. (2013). Annotation and structural elucidation of bovine milk oligosaccharides and determination of novel fucosylated structures. *Glycobiology*, *23*(6), 664–676. https://doi.org/10.1093/glycob/cwt007

Amorim, C., Silvério, S. C., Cardoso, B. B., Alves, J. I., Pereira, M. A., & Rodrigues, L. R. (2020). In vitro fermentation of raffinose to unravel its potential as prebiotic ingredient. *LWT*, *126*, 109322. https://doi.org/10.1016/j.lwt.2020.109322

Bainy, E. M., Tosh, S. M., Corredig, M., Poysa, V., & Woodrow, L. (2008). Varietal differences of carbohydrates in defatted soybean flour and soy protein isolate by-products. *Carbohydrate polymers*, *72*(4), 664–672. https://doi.org/10.1016/j.carbpol.2007.10.008

Barreira, J. C. M., Pereira, J. A., Oliveira, M. B. P. P., & Ferreira, I. C. F. R. (2010). Sugars profiles of different chestnut (*Castanea sativa* Mill.) and almond (*Prunus dulcis*) cultivars by HPLC-RI. *Plant foods for human nutrition (Dordrecht, Netherlands)*, *65*(1), 38–43. https://doi.org/10.1007/s11130-009-0147-7

Bouchard, A., Hofland, G. W., & Witkamp, G.-J. (2007). Properties of sugar, polyol, and polysaccharide water−ethanol solutions. *Journal of Chemical & Engineering Data*, *52*(5), 1838–1842. https://doi.org/10.1021/je700190m

Carter, H. E., McCluer, R. H., & Slifer, E. D. (1956). Lipids of wheat flour. I. characterization of galactosylglycerol components[1]. *Journal of the American Chemical Society*, *78*(15), 3735–3738. https://doi.org/10.1021/ja01596a051

Chappuis, E., Morel-Depeisse, F., Bariohay, B., & Roux, J. (2017). Alpha-Galacto-Oligosaccharides at Low Dose Improve Liver Steatosis in a High-Fat Diet Mouse Model. *Molecules (Basel, Switzerland)*, *22*(10). https://doi.org/10.3390/molecules22101725

Chen, J., Chen, D., Zhang, X., Wang, M., Chen, B., An, D., … Lyu, Q. (2018). Quantification of alcohols, diols and glycerol in fermentation with an instantaneous derivatization using trichloroacetyl isocyanante via liquid chromatography-massspectrometry. *Journal of Chromatography. A*, *1568*, 22–28. https://doi.org/10.1016/j.chroma.2018.07.024

Dai, Z., Feng, S., Liu, A. B., Wang, H., Zeng, X., & Yang, C. S. (2019). Protective effects of α-galacto-oligosaccharides against a high-fat/western-style diet-induced metabolic abnormalities in mice. *Food & function*, *10*(6), 3660–3670. https://doi.org/10.1039/c9fo00463g

Eggert, A., & Karsten, U. (2010). Low molecular weight carbohydrates in red algae – an ecophysiological and biochemical perspective. In J. Seckbach & D. J. Chapman (eds.), *Red algae in the genomic age* (Vol. 13, pp. 443–456). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-90-481-3795-4_24

Fan, P.-H., Zang, M.-T., & Xing, J. (2015). Oligosaccharides composition in eight food legumes species as detected by high-resolution mass spectrometry. *Journal of the science of food and agriculture*, *95*(11), 2228–2236. https://doi.org/10.1002/jsfa.6940

FoodData Central. (2022). U.S. Department of Agriculture FoodData Central. Retrieved April 1, 2022, from https://fdc.nal.usda.gov/index.html

Garg, N., Sethupathy, A., Tuwani, R., Nk, R., Dokania, S., Iyer, A., … Bagler, G. (2018). FlavorDB: a database of flavor molecules. *Nucleic Acids Research*, *46*(D1), D1210–D1216. https://doi.org/10.1093/nar/gkx957

Haselberger, P., & Jacobs, W. A. (2016). Determination of Fructans in Infant, Adult, and Pediatric Nutritional Formulas: Single-Laboratory Validation, First Action 2016.06. *Journal of AOAC International*, *99*(6), 1576–1588. https://doi.org/10.5740/jaoacint.16-0190

Huang, Y.-P., Robinson, R. C., & 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., & 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

Ispiryan, L., Heitmann, M., Hoehnel, A., Zannini, E., & Arendt, E. K. (2019). Optimization and Validation of an HPAEC-PAD Method for the Quantification of FODMAPs in Cereals and Cereal-Based Products. *Journal of Agricultural and Food Chemistry, 67*(15), 4384–4392. https://doi.org/10.1021/acs.jafc.9b00382

Karsten, U., Michalik, D., Michalik, M., & West, J. A. (2005). A new unusual low molecular weight carbohydrate in the red algal genus Hypoglossum (Delesseriaceae, Ceramiales) and its possible function as an osmolyte. *Planta*, *222*(2), 319–326. https://doi.org/10.1007/s00425-005-1527-3

Kitajima, J., Ishikawa, T., & Tanaka, Y. (1998). Water-Soluble Constituents of Fennel. I. Alkyl Glycosides. *Chemical and Pharmaceutical Bulletin*, *46*(10), 1643–1646. https://doi.org/10.1248/cpb.46.1643

Kuo, T. M., VanMiddlesworth, J. F., & Wolf, W. J. (1988). Content of raffinose oligosaccharides and sucrose in various plant seeds. *Journal of Agricultural and Food Chemistry*, *36*(1), 32–36. https://doi.org/10.1021/jf00079a008

Liener, I. E. (1994). Implications of antinutritional components in soybean foods. *Critical reviews in food science and nutrition*, *34*(1), 31–67. https://doi.org/10.1080/10408399409527649

Lignou, S., Parker, J. K., Oruna-Concha, M. J., & Mottram, D. S. (2013). Flavour profiles of three novel acidic varieties of muskmelon (Cucumis melo L.). *Food chemistry*, *139*(1-4), 1152–1160. https://doi.org/10.1016/j.foodchem.2013.01.068

Lu, J., Zhang, Y., Song, B., Zhang, S., Pang, X., Sari, R. N., … Lv, J. (2020). Comparative analysis of oligosaccharides in Guanzhong and Saanen goat milk by using LC-MS/MS. *Carbohydrate polymers*, *235*, 115965. https://doi.org/10.1016/j.carbpol.2020.115965

Massart, D. L., Vandeginste, B. G. M., Buydens, L. M. C., De Jong, S., Lewi, P. J., & Smeyers-Verbeke, J. (1998). *Handbook of Chemometrics and Qualimetrics: Part A*. Elsevier.

Morel, F. B., Dai, Q., Ni, J., Thomas, D., Parnet, P., & Fança-Berthon, P. (2015). α-Galacto-oligosaccharides Dose-Dependently Reduce Appetite and Decrease Inflammation in Overweight Adults. *The Journal of Nutrition*, *145*(9), 2052–2059. https://doi.org/10.3945/jn.114.204909

Obendorf, R. L., Horbowicz, M., Dickerman, A. M., Brenac, P., & Smith, M. E. (1998). Soluble oligosaccharides and galactosyl cyclitols in maturing soybean seeds in planta and in vitro. *Crop science*, *38*(1), 78–84. https://doi.org/10.2135/cropsci1998.0011183X003800010014x

Owolabi, I. O., Dat-arun, P., Takahashi Yupanqui, C., & Wichienchot, S. (2020). Gut microbiota metabolism of functional carbohydrates and phenolic compounds from soaked and germinated purple rice. *Journal of functional foods*, *66*, 103787. https://doi.org/10.1016/j.jff.2020.103787

Pico, J., Martínez, M. M., Martín, M. T., & Gómez, M. (2015). Quantification of sugars in wheat flours with an HPAEC-PAD method. *Food chemistry*, *173*, 674–681. https://doi.org/10.1016/j.foodchem.2014.10.103

Pico, J., Vidal, N. P., Widjaja, L., Falardeau, L., Albino, L., & Martinez, M. M. (2021). Development and assessment of GC/MS and HPAEC/PAD methodologies for the quantification of α-galacto-

oligosaccharides (GOS) in dry beans (Phaseolus vulgaris). *Food chemistry*, *349*, 129151. https://doi.org/10.1016/j.foodchem.2021.129151

Quemener, B., & Brillouet, J.-M. (1983). Ciceritol, a pinitol digalactoside form seeds of chickpea, lentil and white lupin. *Phytochemistry*, *22*(8), 1745–1751. https://doi.org/10.1016/S0031-9422(00)80263-0

Ruiz-Matute, A. I., Sanz, M. L., Moreno-Arribas, M. V., & Martínez-Castro, I. (2009). Identification of free disaccharides and other glycosides in wine. *Journal of Chromatography. A*, *1216*(43), 7296–7300. https://doi.org/10.1016/j.chroma.2009.08.086

Sereewat, P., Suthipinittham, C., Sumathaluk, S., Puttanlek, C., Uttapap, D., & Rungsardthong, V. (2015). Cooking properties and sensory acceptability of spaghetti made from rice flour and defatted soy flour. *LWT - Food Science and Technology*, *60*(2), 1061–1067. https://doi.org/10.1016/j.lwt.2014.10.001

Swennen, K., Courtin, C. M., & Delcour, J. A. (2006). Non-digestible oligosaccharides with prebiotic properties. *Critical reviews in food science and nutrition*, *46*(6), 459–471. https://doi.org/10.1080/10408390500215746

Thodberg, S., Del Cueto, J., Mazzeo, R., Pavan, S., Lotti, C., Dicenta, F., … Sánchez-Pérez, R. (2018). Elucidation of the Amygdalin Pathway Reveals the Metabolic Basis of Bitter and Sweet Almonds (Prunus dulcis). *Plant Physiology*, *178*(3), 1096–1111. https://doi.org/10.1104/pp.18.00922

Wirthensohn, M. G., Chin, W. L., Franks, T. K., Baldock, G., Ford, C. M., & Sedgley, M. (2008). Characterising the flavour phenotypes of almond (*Prunus dulcis* Mill.) kernels. *The Journal of Horticultural Science and Biotechnology*, *83*(4), 462–468. https://doi.org/10.1080/14620316.2008.11512407

Wu, S., Grimm, R., German, J. B., & Lebrilla, C. B. (2011). Annotation and structural analysis of sialylated human milk oligosaccharides. *Journal of Proteome Research*, *10*(2), 856–868. https://doi.org/10.1021/pr101006u

Wunsch, N.-G. (2022, January 17). Forecast of the retail sales of milk alternatives in the United States from 2020 to 2025. Retrieved April 1, 2022, from https://www.statista.com/statistics/1238235/forecast-of-the-retail-sales-of-milk-alternatives/

Wyk, C. J., Kepner, R. E., & Webb, A. D. (1967). Some Volatile Components of Vitis Vinifera Variety White Riesling. 3. Neutral Components Extracted from Wine. *Journal of Food Science, 32*(6), 669–674. https://doi.org/10.1111/j.1365-2621.1967.tb00860.x

Xi, M., Tang, H., Zhang, Y., Ge, W., Chen, Y., & Cui, X. (2021). Microbiome-metabolomic analyses of the impacts of dietary stachyose on fecal microbiota and metabolites in infants intestinal microbiota-associated mice. *Journal of the science of food and agriculture*, *101*(8), 3336–3347. https://doi.org/10.1002/jsfa.10963

Zhang, Y., Su, D., He, J., Dai, Z., Asad, R., Ou, S., & Zeng, X. (2017). Effects of ciceritol from chickpeas on human colonic microflora and the production of short chain fatty acids by in vitro fermentation. *LWT - Food Science and Technology, 79*, 294–299. https://doi.org/10.1016/j.lwt.2017.01.040

Chapter VI

Leveraging bioprocessing strategies to achieve the simultaneous extraction of full-fat chickpea flour macronutrients and enhance protein and carbohydrate functionality

(This chapter was published as a journal article "Machida, K.; **Huang, Y.-P.**; Dias, F.F.G.; Barile, D.; de Moura Bell, J.M.L.N. *Food Bioproc Tech.* 2022, 15, 1760, doi:10.1007/s11947-022-02847-8." Huang, Y.-P. conducted the simple sugar and oligosaccharide analysis and wrote the original draft of the corresponding sections of the manuscript.)

Abstract

The concurrent extraction of lipids, proteins, and carbohydrates can be achieved by aqueous and enzymatic extraction processes, circumventing the low extractability by mechanical pressing and the use of flammable solvents. The use of alkaline protease, preceded or not by carbohydrase pretreatments, was evaluated on the extractability of oil, protein, and carbohydrates from full-fat chickpea flour and protein functionality. Enzymatic extraction increased oil and protein extractability from 49.8 to 72.0–77.1% and 62.8 to 83.5–86.1%, respectively. Although the carbohydrase pretreatments before the addition of protease did not increase oil and protein extractability, the carbohydrate content of the extracts increased from 7.68 to 9.17−9.33 mg/mL, accompanied by the release of new oligosaccharides in the extracts, as revealed by LC-MS/MS characterization. Enzymatic extraction yielded proteins with significantly higher solubility (25.6 vs. 68.2–73.6%) and digestibility (83.8 vs. 90.79–94.67%). Treatment of the extracts with α-galactosidase completely removed the flatulence-causing oligosaccharides (stachyose and raffinose). This study highlights the effectiveness of environmentally-friendly bioprocessing strategies to maximize lipid, protein, and oligosaccharide extractability from full-fat chickpea flour with concurrent improvements in protein solubility and *in vitro* digestibility, reduction of flatulence related oligosaccharides, and generation of a more diverse pool of oligosaccharides for subsequent prebiotic evaluation.

Keywords: Full-fat chickpea flour, protein functionality, enzymatic extraction, aqueous extraction, oligosaccharides.

6.1. Introduction

Due to the increasing world population and the popularity of alternative protein sources, plant-based proteins are becoming the forefront of sustainable food production. Plant-based protein sources provide many benefits including decreased risk of degenerative diseases and reduced environmental impact from its production (González et al. 2011; WHO 2003). Such benefits have promoted increased production and processing of plant-based products, which in turn requires the development of a critical understanding of the impact of key processing conditions (i.e., extraction and recovery) on the extractability and functionality of many plant-based compounds (i.e., proteins, lipids, and carbohydrates).

An area of growing interest is the processing of pulses, which are part of the legume family. Pulses are generally low in fat and high in protein and fiber (Shevkani et al. 2019). Chickpeas, a member of the pulse family, are an example of a good source of carbohydrates (~60 g/100 g), proteins (19 g/100 g), lipids (6 g/100 g), dietary fiber (~17 g/100 g), and other minor constituents (U.S. Department of Agriculture 2019). The increasing popularity and use of chickpeas in the food industry can be explained by its nutritional value and health benefits associated with its consumption (i.e., low glycemic index, prevention of cardiovascular disease, type-2 diabetes) (Wallace et al. 2016). Used in food products worldwide, most notably for hummus production, chickpeas can be a main source of protein in vegan and vegetarian diets (Duranti and Gius 1997). Chickpea protein isolates can also be used not only to improve the nutritional value but the physical and rheological properties of gluten-free food products (Shaabani et al. 2018).

Chickpea proteins, lipids, and carbohydrates can be extracted using numerous methods. The presence of lipids in a food matrix entails the upstream removal of lipids to release proteins and carbohydrates. Traditionally, upstream lipid removal has been accomplished either by solvent

extraction or by the use of mechanical pressing, the selection of which depends on the composition of the material used (De Moura, De Almeida, and Johnson 2009). Despite the environmental and safety issues associated with flammable solvent extraction and low extraction yields associated with mechanical pressing, a protein-rich by-product with varying amounts of residual oil (cake) or compromised functionality (solvent defatted flour) can be obtained (Kim et al. 2021; L'hocine et al. 2006). This sequential approach means that the cake or the defatted flour must be subjected to another processing step to extract proteins and carbohydrates, in addition to removing the remaining lipids using flammable organic solvents.

Alternatively, aqueous extraction processes (AEP) and enzyme-assisted aqueous extraction processes (EAEP) have been used to simultaneously extract lipids, proteins, and carbohydrates from a food matrix without upstream removal of lipids (Campbell and Glatz 2009; De Moura, Maurer, et al. 2011). This environmentally-friendly processing strategy eliminates the negative impact of flammable and hazardous solvents conventionally used for defatting, thanks to the solubilization and transport of proteins to the exterior of the solid matrix, which creates a more porous structure that favors the washing of the oil droplets by the extraction medium (Cheng et al. 2018; Dias et al. 2020). A further improvement upon the AEP process is the enzyme-assisted extraction process (EAEP), which utilizes enzymes such as proteases and carbohydrases to maximize processing extractability. Increased oil and protein extractability in the EAEP has been attributed to enzymatic hydrolysis of the lipid body membrane, proteins, and cell walls (De Moura et al. 2008; Nadar, Pawar, and Rathod 2017).

The successful development of extraction methods for new protein sources depends on the development of fundamental knowledge of the impact of the processing conditions employed (De Moura et al. 2011) on the extractability, composition, and functional properties of the extracted

compounds. Controlled hydrolysis of chickpea protein isolates by immobilized Alcalase has been shown to produce hydrolysates with higher solubility, oil absorption, foaming capacity, and stability (Yust et al. 2010). However, limited emphasis has been given to the development of a holistic understanding of the effects of key extraction parameters (i.e., solids-to-liquid ratio, pH, temperature, incubation time, amount and type of enzyme) on the overall extractability of both lipids and proteins from full-fat chickpea flour and their impact on the functional properties of the extracted proteins.

Because extraction conditions affect yields and the functionality of the target compounds, they play a key role in the processing feasibility and potential applications of the extracted compounds. This work was undertaken to uncover the effects of different enzymatic extraction strategies on the simultaneous extraction of lipids, proteins, and carbohydrates from full-fat chickpea flour and on the functionality of the extracted proteins. Specifically, we evaluated the effectiveness of an upstream enzymatic pretreatment with carbohydrases (cellulase, hemicellulase, and xylanase) before the use of proteases, with respect to lipids and protein extractability, solubility and *in vitro* digestibility of the extracted proteins, and carbohydrate profiling of the extracts. Our working hypothesis was that the use of carbohydrases before the addition of proteases could hydrolyze the cell wall and potentially release new oligosaccharides while also favoring the formation of a more porous structure that could aid in protein solubilization by the aqueous medium, as well as hydrolysis of the protein bodies and oleosin membrane surrounding the lipid bodies by the protease. That could in turn improve the overall process extractability and concurrently produce more soluble and digestible protein hydrolysis products and release a more diverse pool of oligosaccharides with potential health-promoting effects. High-performance anion-

exchange chromatography with pulsed amperometric detection and LC-MS/MS were used to determine the carbohydrate profile of the chickpea extracts.

6.2. Materials and methods

6.2.1. Full-fat chickpea flour and enzymes used in the enzymatic extraction

Commercial Steamed Chickpea flour of the Kabuli variety was kindly provided by Natural Products, Inc (Grinnell, Iowa, USA). Partially dehulled chickpeas (to increase the fiber content of the final product) were steamed to inactivate enzymes and achieve microbial stability before milling (as described by the manufacturer). The chickpea flour contained 7.4 ± 0.1% oil, 25.87 ± 0.07% protein, and 4.69 ± 0.09% moisture, which were determined as described in Section 6.2.3.

The following commercial enzymes were used to assist the enzymatic extraction process (EAEP):

(i) FoodPro Alkaline Protease (also known as Protex 6L) is a bacterial alkaline endoprotease from *Bacillus licheniformis* (pH activity from 8.0 to 10.5, temperature from 45 to 75 ℃, and enzyme activity of 580,000–650,000 DU/g) was provided by the Genencor Division of Danisco (Rochester, NY, USA); (ii) Cellulase from *Trichoderma reesei*, with multiple cellulolytic activities (endo and exo-cellulase, β-glucosidase, β-glucanase, hemicellulose, pectinase, and xylanase) and enzyme activity of 200,000 CU/g at optimal pH from 4.0−6.5 and 45−70 °C was provided by Bio-Cat (Troy, VA, USA); (iii) Hemicellulase from *Aspergillus niger*, with enzyme activity of 600,000 HCU/g and optimal activity at pH 2.0−8.0 and 25−90 °C, was provided by Bio-Cat (Troy, VA, USA); and (iv) Xylanase from *Trichoderma longibrachiatum*, with enzyme activity of 200,000 XU/g and optimal activity at pH 3.5-6.5 and 40-70 °C, was provided by Bio-Cat (Troy, VA, USA).

6.2.2. Tailoring enzyme use to maximize the simultaneous extraction of lipids, proteins, and carbohydrates from full-fat chickpea flour

The effect of using protease (EAEP), alone or in combination with different carbohydrase pretreatments, was evaluated on the extractability of proteins, lipids, and carbohydrates from chickpea flour (Figure 6.1). A non-enzymatic aqueous treatment (AEP) was used as the control.

The AEP (control, no enzyme use) was carried out by dispersing 50 g of chickpea flour into 500 mL of water to achieve a 1:10 solids-to-liquid ratio (SLR). The slurry pH was adjusted to pH 9.0 to favor protein solubility and extractability (Almeida et al. 2019) and kept at 50 °C under constant stirring for 60 min. For the EAEP, the potential benefits of using an upstream treatment with carbohydrases, before the alkaline protease addition, were evaluated. The following enzymatic strategies were evaluated: EAEP 1: 0.5% (w/w) of alkaline protease at pH 9.0 for 60 min; EAEP 2: 0.5% (w/w) of carbohydrases (0.25% of cellulase + 0.25% of hemicellulase) at pH 6.0 for 30 min followed by the addition of 0.5% of alkaline protease (w/w) at pH 9.0 for 60 min; and EAEP 3: 0.5% (w/w) of carbohydrases (0.17% of cellulase + 0.17% of hemicellulase + 0.17% of xylanase) at pH 6.0 for 30 min followed by the addition of 0.5% of alkaline protease (w/w) at pH 9.0 for 60 min. For the EAEP, extractions were performed at the same SLR and temperature as the AEP, and pH conditions were selected based on the enzyme manufacturer's recommendations.

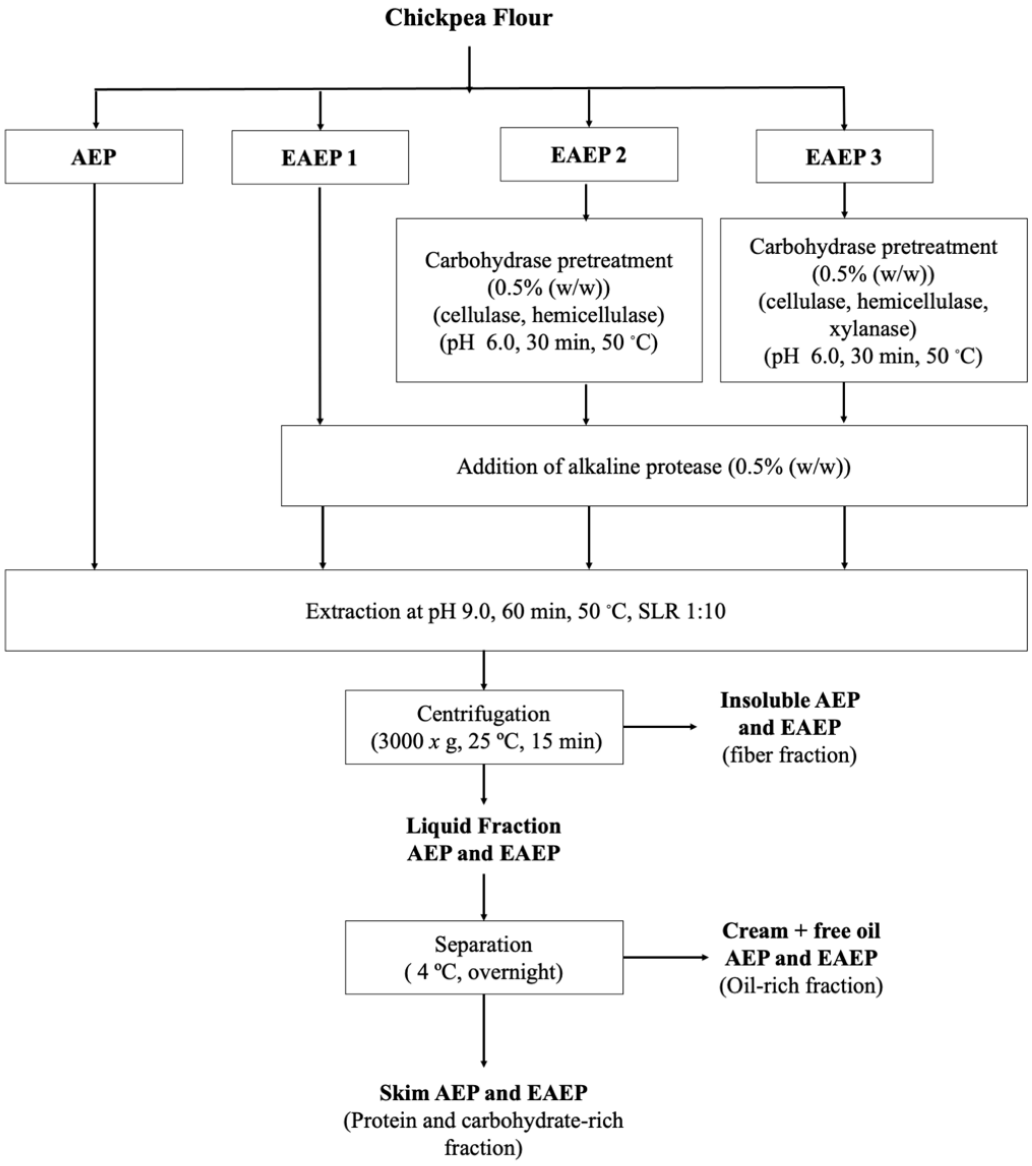


Figure 6.1. Process flow diagram for each extraction treatment. AEP: no enzyme; EAEP 1: 0.5% protease; EAEP2: 0.25% of cellulase and 0.25% of hemicellulase (total of 0.5% of carbohydrases); EAEP3: 0.17% of cellulase, 0.17% of hemicellulase and 0.17% of xylanase (total of 0.5% of carbohydrases). w/w: weight of enzyme/weight of flour.

After extracting, the resulting slurry was centrifuged at 3000 *x g* for 30 min at 4 °C to separate the insoluble fraction (containing the unextracted compounds) from the liquid phase (containing the extracted compounds). The liquid fraction was placed in a separatory funnel and allowed to settle overnight at 4 °C to separate the oil-rich fraction (cream and free oil) from the protein- and carbohydrate-rich fraction (skim). Each extraction condition was carried out in triplicate.

Chickpea full-fat flour (starting material for the extraction) and all fractions generated by the AEP and EAEP were characterized for oil, protein, and carbohydrate contents (as described in sections 6.2.3 and 6.2.7). Total oil extraction yield (TOE), oil distribution in the fractions (free oil yield, oil yield in the cream, skim, and insoluble), total protein extraction yield (TPE), and protein distribution in the fractions (protein yield in the cream, skim, and insoluble) were determined according to Equations (1 and 2), respectively (Souza et al., 2019):

$$TOE/TPE\ (\%) = [100\ - (\frac{Oil/Protein\ (g)\ in\ Insoluble}{Oil/Protein\ (g)\ in\ the\ chickpea\ flour})\ ] \times\ 100 \qquad (1)$$

$$Oil/Protein\ distribution\ in\ the\ fractions\ (\%) = (\frac{Oil/Protein\ (g)\ in\ fraction*}{Oil/Protein\ (g)\ in\ the\ chickpea\ flour}) \times\ 100 \quad (2)$$

*The fractions relate to free oil, cream, skim, and insoluble.

6.2.3. Proximate analysis

Oil (acid hydrolysis - AOCS method 989.05), dry matter (AOCS method 925.09) and protein content (AOAC 992.23- Dumas combustion method, 6.25 of nitrogen conversion factor) were determined in the starting material and fractions generated in the extraction. Analyses were performed in duplicate for each extraction (n=6).

6.2.4. Protein degree of hydrolysis of skim fractions

The protein degree of hydrolysis (DH) of AEP and EAEP skim fractions was determined by the o-phthaldialdehyde (OPA) method (Nielsen et al. 2001) in a spectrophotometer at 340 nm using L-serine solution as the standard. The DH was determined as the ratio of h (number of hydrolyzed bonds) and htot (total number of peptide bonds per protein equivalent - 7.22 for chickpeas (Kou et al. 2013).

6.2.5. Low molecular weight (MW) polypeptide profile characterization of AEP and EAEP skim proteins by sodium dodecyl sulfate-polyacrylamide (SDS-PAGE)

SDS-PAGE of the skim fractions was performed using a 12% acrylamide gel and 30 µg of protein/ well. A low MW range standard ladder (14.4–97.4 kDa) (Bio-Rad, Hercules, CA, USA) was used. Relative quantification and polypeptide distribution were performed using an Imager system and Image Lab software (Bio-Rad, Hercules, CA, USA).

6.2.6. Solubility of skim proteins

Approximately 15 mL of AEP and EAEP skims from each extraction replicate (n=3) were freeze-dried on a FreeZone 4.5 Liter Benchtop Freeze Dry System (Labconco, Kansas City, MO, USA) and stored at −20 ◦C for subsequent solubility tests. Protein solubility of freeze-dried AEP and EAEP skim proteins was evaluated by preparing a 10 mL of a 1% (w/v) skim solution in a 30 mL beaker and adjusting the pH of the protein solution to 4.0 and 9.0 by the addition of 1 M HCl or 1 M NaOH solution. Solutions were vigorously mixed at 150 rpm for 1 hour at room temperature and then centrifuged at 10,000 x g at 20 °C for 10 min. The total protein content was determined as described in item 2.3. All samples were analyzed in duplicate. Skim protein solubility was expressed as the percentage ratio of the supernatant protein content to the sample protein content.

6.2.7. Carbohydrate profile, quantification, and α-galactosidase treatment of AEP and EAEP skim fractions

The use of carbohydrases during the extraction process and as a post-extraction strategy can result in the production of chickpea extracts containing a diverse oligosaccharide profile while eliminating flatulence promoting oligosaccharides such as raffinose and stachyose, which are present in high amounts in pulses, including chickpeas. The effects of the use of different carbohydrases during the extraction and post-extraction on the content and profile of carbohydrates of the skim fractions were evaluated by different analytic techniques.

6.2.7.1. Quantification of total carbohydrates by spectrophotometry

The total carbohydrate content of the skim fractions was determined using the Phenol-sulfuric method (Masuko et al. 2005). 15 µL of the sample along with 15 µL of nanopure water were added to the well. The well plate was shaken at 300 rpm for 1 min then 150 µL of 98% sulfuric acid was added to each well. The microplate was then incubated at 85 °C for 15 minutes in an incubating thermal shaker (Thermalshake, VWR, Radnor, Pennsylvania, USA), followed by the addition of 30 µL of 5% (w/v) phenol/water solution. A calibration curve made using glucose as a standard (from 4 to 20 µg, R^2= 0.9937) was used to quantify the total carbohydrates. After vigorous mixing, the samples were measured using a microplate reader (SpectraMax iD5, Molecular Devices, San Jose, CA) at 490 nm.

6.2.7.2. Quantification of oligosaccharides and simple sugars by high-performance anion-exchange chromatography

Soluble carbohydrate profiles of skims were quantified by high-performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD, Dionex ICS-5000+, Thermo Fisher Scientific, Sunnyvale, CA, USA). Aliquots of 200 µL of samples were

mixed with 400 μL of ethanol in 1.5 mL tubes, vortexed, and incubated at -30 °C for 1 h. After being centrifuged at 4 °C for 30 min (13,000 × g), the supernatant was dried under vacuum (MiVac Quattro concentrator, Genevac Ltd., Ipswitch, UK). Samples were diluted as appropriate and filtered through a 0.2 μm syringe filter into 1.5 mL vials with septa. Glucose, galactose, and fructose were separated on a CarboPac PA10 column (4 × 250 mm) with a CarboPac PA10 guard column (4 × 50 mm) at a flow rate of 1.0 mL/min. The mobile phase was maintained at an isocratic condition of 10 mM NaOH for 12 min and was increased to 100 mM NaOH in 13 min. Sucrose, raffinose, and stachyose were separated on a CarboPac PA200 column (3 × 250 mm) with a CarboPac PA200 guard column (3 × 50 mm) by isocratic elution using a mobile phase of 50 mM NaOH at a flow rate of 0.5 mL/min. Both columns were washed with 200 mM NaOH for 5 min after each run and equilibrated with the respective initial mobile phases for 10 min before the next injection. Calibration curves were built by using 1–60 μg/mL of glucose, galactose, and fructose, and 0.1–10 μg/mL of sucrose, raffinose, and stachyose. An analytical replicate was conducted for each replicate of extraction and α-galactosidase treatment (n=3).

6.2.7.3. α-galactosidase treatment of raffinose and stachyose in the skim fractions

Because of the presence of flatulence promoting oligosaccharides in the skim fractions, an α-galactosidase treatment was used to reduce the concentration of stachyose and raffinose in the skim fractions. Because the amount of stachyose and raffinose was not statistically different within the enzymatic treatments, the EAEP 1 skim was selected to demonstrate the effectiveness of the α-galactosidase treatment. 1000 μL of EAEP 1 skim was adjusted to pH 6 with 1 M HCl, and α-galactosidase (Bio-Cat, Inc., Troy, VA, USA) was added to achieve a 0.25 % (w/v) concentration. The EAEP 1 skim was incubated at 40 °C for 0, 15, 30, and 60 min at 90 rpm in a water bath. Skim samples were placed in an ice bath to stop the reaction and stored at 4 °C until analyzed. The

quantification of simple sugar (glucose, galactose, fructose, and sucrose) and oligosaccharide (raffinose and stachyose) was carried out by HPAEC-PAD as described in section 6.2.7.2.

6.2.7.4 Mass spectrometry characterization of oligosaccharides in the extracts

The oligosaccharide profile of the skims was characterized by liquid chromatography-tandem mass spectrometry (LC-MS/MS). The reconstituted supernatant fractions obtained from ethanol precipitation (described in section 6.2.7.2) were further purified by solid-phase extraction (SPE). Mixed-mode SPE cartridges, which retain compounds by both hydrophobic interaction and strong cation exchange, were used for separating oligosaccharides from peptides (Huang et al., 2022). The reconstituted samples (150 μL, equivalent to 30 μL of skims) were premixed with 150 μL 0.2% formic acid and then loaded to Strata-X-C SPE cartridges (30 mg/1 mL, Phenomenex, Torrance, CA, USA) preconditioned with acetonitrile and 0.1% formic acid. Oligosaccharides were eluted with 3 mL 0.1% formic acid and further loaded to porous graphitic carbon SPE microplate (Glygen, Columbia, MD, USA) preconditioned with 80% acetonitrile with 0.1% trifluoroacetic acid and water. The microplate wells were washed with water for eliminating salts and flushed sequentially with 40% acetonitrile (fraction 1) and 40% acetonitrile with 0.1% trifluoroacetic acid (fraction 2) for eluting oligosaccharides. The collected oligosaccharide fractions were dried in a centrifugal evaporator. Fractions 1 and 2 were combined after dissolving the dried samples in water. For oligosaccharide characterization, the combined samples of the three replicates of extraction were pooled and injected into the LC-MS/MS (one injection for each treatment). For relative quantification, one injection was made for each extraction replicate.

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, USA) equipped with an Agilent PGC-Chip II (porous graphitized carbon chip with a 40 nL enrichment column and a 75

μm × 43 mm analytical column). The capillary pump delivered 3% acetonitrile with 0.1% formic acid (v/v/v) at a flow rate of 4 μL min^{-1} and loaded samples into the enrichment column. The injection volume was 2 μL for each sample. The nano pump delivered mobile phase composed of 3% acetonitrile with 0.1% formic acid (v/v/v) (solvent A) and 89.9% acetonitrile with 0.1% formic acid (v/v/v) (solvent B). The analytes were separated at a flow rate of 0.3 μL min^{-1} with 0% B from 0.0–2.5 min; 0–16% B from 2.5–20.0 min; 16–44% B from 20.0–30.0 min; 44–100% B from 30.0–35.0 min; 100% B 35.0–45.0 min. The mobile phase was switched to 100% A and equilibrated for 15 min before the next injection. The capillary voltage was set at 1850 V to maintain a stable spray. The drying gas was set at 350 °C at a flow rate of 5 L min^{-1}. The scanning mass range was *m/z* 150–2500 for MS and 50–2500 for MS/MS. Collision energy for tandem MS was set by a formula of [0.02 × (*m/z*) − 3.5]. Data analysis was conducted in MassHunter Qualitative Analysis B.07.00 (Agilent Technologies). Oligosaccharides were identified by inspecting fragmentation patterns in tandem MS spectra. For relative quantification, peak areas of the identified oligosaccharides were integrated from merged extracted ion chromatograms, including the precursor ions and corresponding in-source fragment ions, to approach the real relative quantities (Huang et al., 2022b).

6.2.8. In vitro skim protein digestibility

Protein digestibility of AEP and EAEP skim proteins was measured as described by (Bornhorst and Singh 2013; de Souza et al. 2020). Five mL of liquid skim fractions were mixed with 3.33 mL of SSF (Simulated Saliva Fluid) and vortexed. Subsequently, 6.66 mL of SGF (Simulated Gastric Fluid) was added. Afterward, the pH was adjusted to 3.0 and the samples were placed into a water bath (37 °C, 140 rpm, 2 h). Then, 10 mL of SIF (Simulated Intestinal Fluid) was added, and the pH was adjusted to 7.0. The samples were incubated in a water bath at 37 °C,

140 rpm, for 2 h. To stop the digestion, samples were heated in a water bath at 85 °C for 3 min. Trichloroacetic acid (TCA) was added in a 1:1 (v/v) proportion to the samples to achieve a final 12% (w/w) TCA concentration. The samples were centrifuged at 3578 × *g* for 30 min at 4 °C. The precipitate, protein nitrogen fraction (PN), was analyzed for the protein content. A before digestion control with sample and water, instead of simulated liquids, was performed and an enzyme blank with water, instead of the sample, was also performed. The digestibility was calculated as described by (Zhong et al. 2012).

$$Digestibility\ (\%) = \frac{PN_{before} - (PN_{after} - PN_{enzyme\ blank})}{PN_{before}} \quad (3)$$

Where PN_{before} = protein before digestion, PN_{after} = protein after digestion, $PN_{enzyme\ blank}$ = enzyme blank. The PN (protein nitrogen fraction) was measured in the samples by the Dumas method using a conversion factor of 6.25 (Vario MAX cube, HE, DE) before and after the digestion.

6.2.9. Statistical analysis

Extractions were performed in triplicate and the functional analyses were performed in duplicate. The results were expressed as the mean ± standard deviation (SD) of the replicates. Replicates of each measurement were analyzed by ANOVA with generalized linear models from the Statistica software (version 13.5.0.17 1984-2018, TIBCO Software Inc, Palo Alto, CA, USA). Multiple comparisons of least-square means were made by Tukey's adjustment with the level of significance set at $p < 0.05$. Statistical significant differences were denoted by different letters, with the letter "a" being assigned to the highest value.

6.3. Results and Discussion

6.3.1. Effects of extraction conditions on oil and protein extraction yields

The use of selected enzymes to assist the extraction of plant-based matrices has been successfully used as an effective strategy not only to increase the extractability of desired compounds (i.e., lipids, proteins, carbohydrates, phenolics) but to impart structural modifications in the food matrix that can lead to the production of compounds with desired functional and biological properties (i.e., higher protein digestibility and solubility, release of prebiotic oligosaccharides and antioxidants, among others (de Souza et al., 2020; Dias and Bell, 2022)).

The effectiveness of using alkaline protease, preceded or not by the use of selected carbohydrases, on the extractability of lipids and proteins from full-fat chickpea flour is shown in Figures 2A and 2B. Enzymatic extraction significantly increased the overall extractability of lipids from full-fat chickpea flour (Figure 6.2A) compared with the control (AEP, no enzyme use). When not using enzymes (AEP), 49.78 ± 2.08% of the available oil in the chickpea flour was extracted. However, oil extraction yields increased to 77.15 ± 5.87% for the EAEP 1 (using only protease), followed by 73.45 ± 1.54% for the EAEP 2 (cellulase + hemicellulase pre-treatment followed by protease) and 72.02 ± 1.19% for the EAEP 3 (cellulase + hemicellulase + xylanase pretreatment followed by protease). The higher oil extraction yields observed for EAEP treatments can be primarily attributed to the modes of action and effectiveness of the protease used. Proteases can hydrolyze the oleosin membrane of the lipid bodies, releasing free oil into the aqueous medium (Campbell et al. 2011). In addition, protein removal from the matrix by solubilization or proteolysis leaves behind a more porous structure that facilitates the release of the oil. On the other hand, the AEP relies primarily on the solubilization of the proteins into the aqueous medium, without the benefit of proteolysis above described. Therefore, lipids are solely extracted through

washing out of the matrix. Despite the higher EAEP oil extraction yields, compared with the AEP, oil extraction yields were not statistically different among the enzymatic treatments. Additional carbohydrase pretreatments did not significantly increase lipid extractability, therefore not justifying the additional use of enzyme, energy, and time. When looking at the oil distribution for the EAEP 1, although 77.15% of the chickpea flour oil was extracted, only 0.43% of the total extracted oil was present as free oil, while 42 and 57% of the extracted oil were present in the cream (oil-rich emulsion) and skim fractions, respectively. Comparatively, for the AEP, only 0.15% of the extracted oil was present as free oil, with 15 and 32% being present in the cream and skim, respectively. While the amount of free oil extracted by the AEP and EAEP was not statically different (0.14 vs. 0.24–0.29% yield), the use of enzyme in EAEP 1 and 2 significantly increased the oil yield in the cream (16.06 vs. 32.23–32.56% yield). Since there are no methods available to recover the diluted oil from the skim fraction, shifting more lipids into the cream fraction is of key importance to favor the overall recovery of the extracted oil, which entails the development of additional demulsification studies to breakdown the cream emulsion (De Moura and Johnson 2009), which is beyond the scope of this work. Our results are in agreement with the literature (Dias et al. 2020; Souza et al. 2019), which demonstrates that most of the oil extracted through the AEP and EAEP is entrapped in the cream fraction.

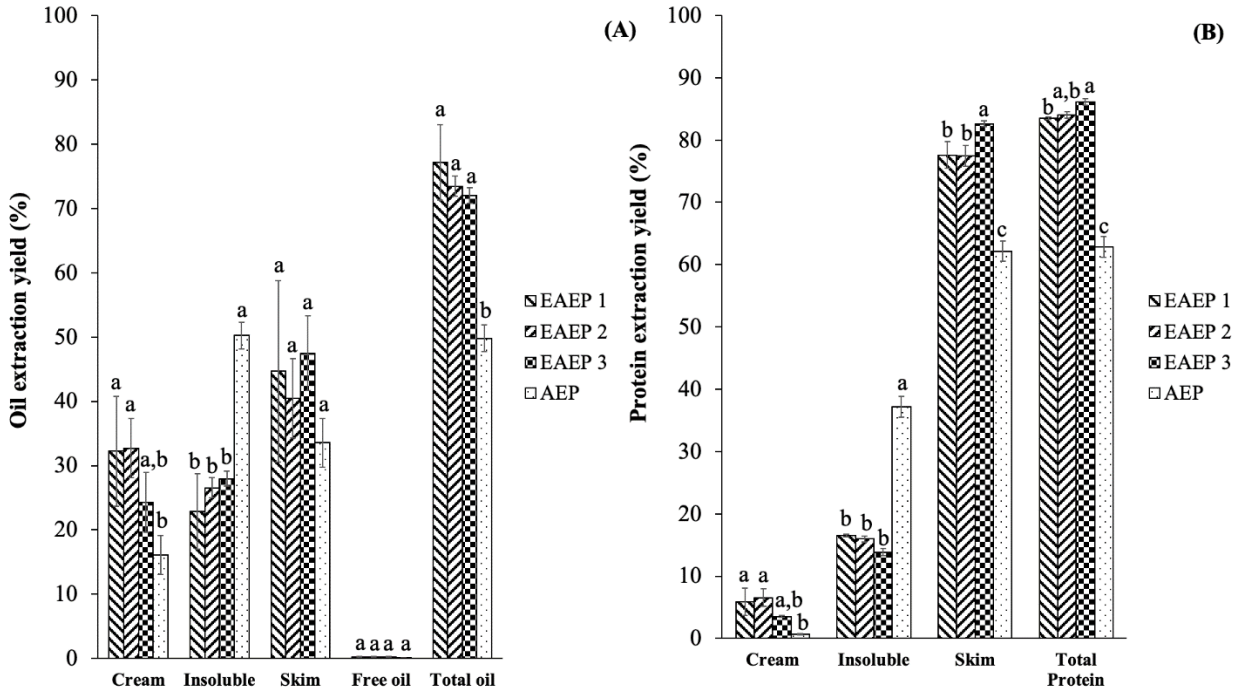


Figure 6.2. Oil (A) and protein (B) extraction yields and distribution in the fractions. Different letters indicate statistically significant differences by one-way ANOVA followed by Tukey's test at $p<0.05$.

Although the lipid content in AEP and EAEP skim fractions is low (~0.3 and 0.4%, respectively), the high volume of skim produced accounts for a significant portion of the oil in the chickpea flour (up to 47.5% yield). On a dry basis, the oil content of AEP and EAEP skims are 7.77 and 7.91%, respectively. When comparing the oil distribution with previous studies for other food matrices, lower oil yields have been reported for skim fractions produced by enzymatic hydrolysis (Protex 6L) of extruded soybean flakes (14% yield, De Moura et al. 2008) and almond cake (14% yield, and Souza et al. 2019). This could be attributed to differences in the composition (i.e., lipids, protein) of the starting materials and processing conditions used (i.e., milling, flaking, extruding, type of enzyme). As an example, the initial oil content in the chickpea flour is very low (7%) compared to that of soybeans (21%) and almonds cake (16.25%). Because there are no

methods available to recover oil from the skim fraction, and lipids can reduce the skim protein solubility (Almeida et al. 2019), it is important to identify processing conditions leading to reduced oil in the skim fraction, which should, in turn, increase the lipid content in the cream fraction for subsequent recovery as free oil.

Overall, the addition of a carbohydrase pretreatment in EAEP 2 and 3 did not significantly increase oil extractability compared with the use of protease alone (EAEP 1), nor altered the distribution of the extracted oil among the fractions. However, the use of protease in all enzymatic treatments significantly increased oil extractability and oil yield in the cream when compared with the AEP.

Figure 6.2B shows the significant increase in protein extractability when enzymes were used to assist the extraction (EAEP) compared with the control (AEP). Enzymatic extraction significantly increased protein extractability from 62.81 ± 1.68% (AEP) to 83.49–86.13% (EAEP). However, extraction yields within the enzymatic strategies evaluated were very similar (83.49 ± 0.19% for EAEP 1, 84.04 ± 0.49% for EAEP 2, and 86.13 ± 1.51% for EAEP 3). The small increase in protein extractability observed for EAEP 3, compared with EAEP 1, could be attributed to the carbohydrase pretreatment applied before the addition of the protease, indicating the breakdown of the cell walls by the carbohydrases and the additional extraction time (30 min) helped with the additional release of proteins from the chickpea flour. Nonetheless, considering the additional use of 0.5% of enzyme and additional reaction time (30 min) when performing the carbohydrase pretreatment, the modest increment in protein extractability observed compared with the use of the protease alone (83.5 vs. 84.0–86.0%) does not justify the inclusion of the additional pretreatment. As expected, the higher protein extractability observed for EAEP treatments led to the production of skim fractions with higher yields (77.49–82.62%) compared with the AEP (62.14%). From the

83.5–86% protein extracted, 77–83% and 61% of the extracted protein was present in the EAEP and AEP skims, respectively. The higher extractability of the EAEP was reflected by the higher protein content of the EAEP skims (2.34-2.36%) compared with the one from the AEP skim (1.83%).

Importantly, the distribution of extracted proteins was influenced by the different modes of action of the enzymes used in the EAEP treatments. While the AEP produced a skim fraction with the lowest protein yield (62.14%), the use of cellulase + hemicellulase + xylanase before the addition of the protease (EAEP 3) led to higher protein yield in the skim (82.62%) for subsequent recovery, compared with the use of cellulase + hemicellulase before the use of protease (EAEP 2) (77.49%) or protease alone (EAEP 1) (77.61%). A similar trend was observed for the cream fraction, where EAEP 1 and 2 led to the production of a cream fraction with a higher protein yield (5.88 and 6.56%, respectively) compared with the AEP (0.67%).

To the best of our knowledge, there are no reports describing the effectiveness of aqueous and enzymatic extraction processes to simultaneously extract lipids and proteins from full-fat chickpea flour, which hinders the comparison of our data with the literature. Our findings are consistent with the ones presented for AEP and EAEP of other food matrices. De Moura et al. 2008 reported protein extraction yields of 85% when using Protex 6L to assist the extraction of extruded soybean flakes and Souza et al. 2019 reported an increase in protein extractability from 69.6% (AEP) to 75% when Protex 6L was used to assist the extraction from the almond cake.

6.3.2 Effects of extraction conditions on the degree of hydrolysis and low MW polypeptide profile of AEP and EAEP skim proteins

During proteolysis, the breakdown of peptide bonds results in an increased concentration of primary amines, corresponding to an increase in the degree of hydrolysis (DH). Because the DH

often has a significant impact on the functional properties of the extracted proteins (Ghribi, Maklouf Gafsi, et al. 2015), understanding the effects of extraction conditions on the DH and protein functionality becomes necessary to further identify possible industrial applications for the extracted proteins.

Enzymatic extraction significantly increased the DH from 10.0% (AEP) to 23.3, 25.0, and 25.5% for the EAEP 2, EAEP 1, and EAEP 3, respectively (Figure 6.3). No significant difference was observed for the DH amongst the enzymatic treatments, in agreement with the use of the same amount of protease in all EAEP treatments. Our results are in agreement with the literature where the use of enzymes to assist the extraction leads to a higher DH (Ghribi, Sila, et al. 2015).

The protein profile of chickpea skim proteins is shown in Figure 6.3. AEP skim proteins (unhydrolyzed proteins) presented a band at ~66 kDa that could be attributed to convicilin, a protein with molecular weight between 68 to 70 kDa (Tzitzikas et al. 2006), corresponding to 18.1% of the protein in the lane. Another intense band can be seen at 45−47 kDa, which might correspond to the vicilin protein, which has three different polypeptide subunits with molecular weights of 53, 47, and 43 kDa (Romero et al. 1975). The major bands observed at ~40 and 20 kDa can be attributed to the acidic (α) subunit of legumins and the basic (β) subunit of legumins, respectively (Boulter and Croy 1997). Moreover, the bands at 37 kDa and 27 kDa could indicate the presence of lectins (Sathe 2002). The legumin alpha-subunit and the lectins correspond to 12.3% of the protein in the lane. Our results agree with the ones reported by Chang et al. 2012, which reported globulin protein 11S legumins and 7S vicilins as the major protein fractions and 2S albumin as the minor protein fraction in chickpea flour.

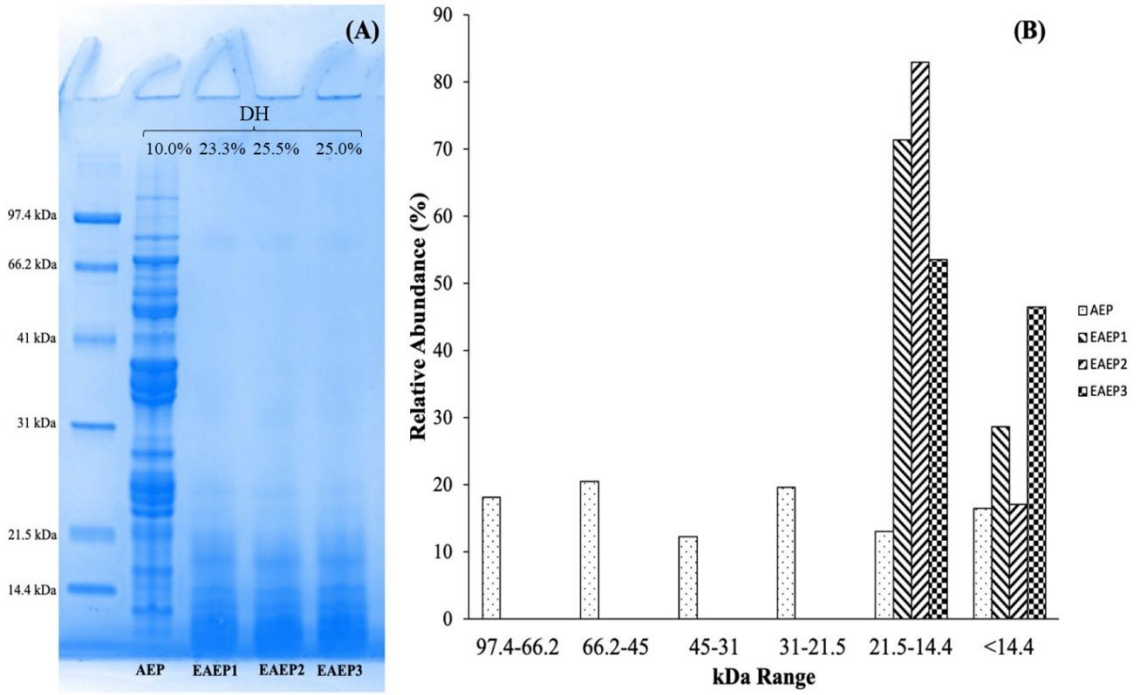


Figure 6.3. Sodium dodecyl sulfate-polyacrylamide 12% gel electrophoresis profiles (A) of AEP and EAEP skim proteins/peptides and degree of hydrolysis (DH) and (B) molecular weight distribution of AEP and EAEP skim protein bodies.

The use of enzymes to assist the extraction (EAEP 1–3) promoted the complete hydrolysis of proteins with MW > 21 kDa, indicating total hydrolysis of convicilin, vicilin, and legumin α-subunit, and partial hydrolysis of the legumin β-subunit, in congruence with the significantly higher degree of hydrolysis of those samples. Moreover, a significant increase in the relative abundance of protein hydrolysis products and peptides with MW < 14 kDa can be observed for most EAEP samples. Ghribi et al. 2015 showed a significant decrease in the ~45–66 kDa and ~34–45 kDa molecular weight bands due to increasing enzymatic hydrolysis of chickpea proteins.

6.3.3. Effects of extraction conditions on protein solubility

Solubility is an important functional property of proteins because of its impact on food applications. Soluble proteins can be integrated into food products whose pH can vary widely, while insoluble proteins may be limited in their application and therefore its desirability. Chickpea protein peptides are needed to be functional, and specifically soluble, to enhance their applications in the food industry. Boye et al. 2010 reported that unhydrolyzed extracted chickpea proteins had higher water and oil absorption capacities, and emulsifying capabilities than other pulse proteins while having similar solubility and gelation capabilities. As the result of proteolysis, smaller peptides are released which can be significantly more soluble than larger protein bodies (Carbonaro et al. 1997).

Because enzymatic hydrolysis can significantly affect protein functionality, we evaluated the impact of the AEP and EAEP on the solubility of extracted proteins at pH 4.0 (which is close to the isoelectric point of chickpea proteins (4.3, (Sánchez-Vioque et al. 1999) and pH 9.0 (Figure 6.4A and B).

At pH 4, where chickpea protein solubility is unfavored by the proximity to its isoelectric point (pI), enzymatic extraction significantly improved protein solubility (25.6% AEP vs. 68.2–73.6% EAEP) (Figure 6.4A). However, no statistically significant differences were observed amongst the enzymatic treatments. These results demonstrate that the use of enzymes during the extraction can indeed generate smaller and more soluble peptides, in agreement with previous studies reporting the beneficial effects of proteolysis on the solubility of almond proteins (Almeida et al. 2019; Souza et al. 2019). Increased solubility of EAEP skim proteins at pH 4.0, compared with AEP skim proteins at the same pH, agrees with the higher DH of EAEP skim proteins. However, at pH 9.0, AEP and EAEP skim proteins exhibited similar high solubility, with values

ranging from 85 to 88% (Figure 6.4B). Higher solubility of AEP and EAEP skim proteins at pH 9.0 is attributed to a higher negative net charge of the proteins, which enhances electrostatic repulsion between protein molecules thus favoring its solubility. Conversely, at the isoelectric point, the net-zero charge of the proteins enhances the attractive forces within the protein molecules, which in turn reduces their solubility in the aqueous medium (Zayas 1997). It is not surprising that chickpea protein solubility at acidic pH, which is near the isoelectric point, is lower than that at alkaline pH. In that view, all enzymatic treatments significantly increased protein solubility at acidic pH. Such increase in solubility has been attributed to an enhanced net charge of the hydrolysates, which can heighten molecular electrostatic repulsion, thus favoring the unfolding of proteins and increasing protein-water interactions (Ghribi et al., 2015). Increased protein solubility at acidic pH is of particular importance as it can open up potential uses of the hydrolysates in specific industrial food formulations involving acidic pH (e.g., protein-rich beverages, protein supplements).

It is important to highlight the potential impact of enzymatic extraction on the functional properties (e.g., foaming, gelling, and emulsification properties) of the extracted protein. For instance, enzymatic extraction of almond cake proteins resulted in hydrolyzates with reduced foaming and emulsification properties, suggesting that extensive hydrolysis (DH>10%) can reduce some functional properties (Souza et al., 2020). Similar results were reported for chickpea protein hydrolysates, with higher DH resulting in hydrolyzates with reduced emulsification properties (Ghribi et al., 2015). However, moderate hydrolysis during the enzymatic extraction of almond flour (DH~7) resulted in the production of hydrolyzates with higher emulsifying properties and foaming capacity at pH values close to the protein isoelectric point (Dias and Bell, 2022). Therefore, a holistic evaluation of the impact of extraction conditions on the protein structure and

functionality, which depends on the matrix characteristics and upstream unit operations employed, is necessary to identify potential applications of the extracted protein and re-evaluate the selection of the extraction conditions.

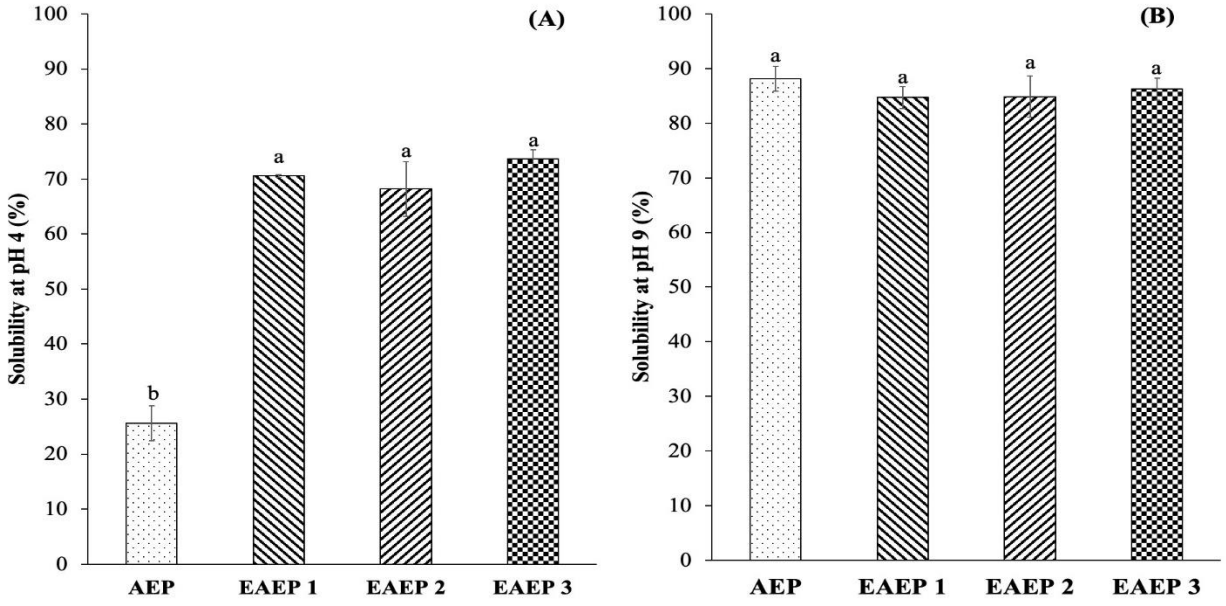


Figure 6.4. AEP and EAEP skim protein solubility at pH 4 (A) and pH 9 (B). Different letters indicate statistically significant differences by one-way ANOVA followed by Tukey's test at $p<0.05$.

6.3.4. Effects of extraction conditions on carbohydrate content, profile, and a-galactosidase treatment of AEP and EAEP skim fractions

6.3.4.1. Quantification of oligosaccharides and free monosaccharides sugars by spectrophotometry and high-performance anion-exchange chromatography

In addition to being a source of raffinose, stachyose, and verbascose, chickpeas are a source of dietary fiber (18–22 g per 100 g of flour), from which 10–18 g is comprised of insoluble fiber and 4–8 g is comprised of soluble flour (Tosh and Yada 2010). The use of carbohydrase pretreatments, before proteolysis, was evaluated as a strategy to improve the bio-functionality of

the protein extracts through the release of potentially prebiotic oligosaccharides into the skim fractions via the breakdown of the cell wall polysaccharides. The effects of the extraction methods used on the total carbohydrate content of the skim fractions and oligosaccharides which are known to exist in chickpeas and for which high-purity standards exist were evaluated by two assays (Table 6.1).

Table 6.1. Total carbohydrates (measured by spectrophotometry), and raffinose, stachyose, sucrose, and monosaccharide concentrations (mg/mL) of AEP and EAEP skims (measured by HPAEC-PAD). Monosaccharides include glucose, galactose, and fructose.

	Total carbohydrates	*Raffinose*	*Stachyose*	*Sucrose*	*Free monosaccharides*
AEP	7.68±0.60[c]	0.88±0.03[a]	2.70±0.15[a]	4.70±0.23[a]	trace
EAEP 1	8.37±0.51[b,c]	0.88±0.03[a]	2.75±0.05[a]	4.70±0.15[a]	trace
EAEP 2	9.17±0.52[a,b]	0.83±0.16[a]	2.53±0.43[a]	4.28±0.64[a]	trace
EAEP 3	9.33±0.29[a]	0.97±0.03[a]	2.80±0.00[a]	4.68±0.03[a]	trace

[1] Different letters in the same column indicate statistically significant difference by one-way ANOVA followed by Tukey's test at $p<0.05$.

Indeed, the enzymatic treatments significantly increased the total carbohydrate content of the skim fractions from 7.68 mg/mL (AEP) to 8.37−9.33 mg/mL (EAEP1–3) (Table 6.1). The higher carbohydrate content of EAEP 2 and 3 skim fractions can be attributed to the use of cellulase and hemicellulase in the EAEP 2 and cellulase, hemicellulase, and xylanase in EAEP 3, which likely promoted the breakdown of the cell wall cellulose and hemicellulose (Reese et al. 1950) into smaller carbohydrate structures. While the EAEP increased the overall extractability of chickpea carbohydrates, the amount of sucrose (4.28–4.70 mg/mL) and major oligosaccharides raffinose (0.83–0.97 mg/mL) and stachyose (2.53–2.80 mg/mL) in the skim fractions was not statistically different within the extraction processes evaluated (AEP vs. EAEP 1–3). This is not surprising because the enzymes used in the EAEP 1–3, including alkaline protease, cellulase, hemicellulase,

and xylanase, do not target glycosidic linkages in sucrose, stachyose, and raffinose to cause their degradation. Sucrose, stachyose, and raffinose were also not expected to be generated, under the action of the carbohydrases, since they are not part of cell wall polysaccharides' structures. Besides, possibly due to the small size of sucrose, stachyose, and raffinose, their extractability was already high in the AEP and did not further increase when the alkaline protease and carbohydrases were used (EAEP 1–3). Because the increment in the total carbohydrate content in the EAEP skims was not associated with the release of free monosaccharides (all in trace concentration) nor with an increase in sucrose and major oligosaccharides such as raffinose or stachyose, LC-MS/MS was used to evaluate the potential release of oligosaccharide by the enzymatic treatments and characterize the composition of the newly generated oligosaccharides.

6.3.4.2. α-Galactosidase treatment of raffinose and stachyose in the skim fractions

Chickpeas are rich in raffinose, stachyose, and verbascose, whose simplicity of monosaccharide composition (galactose, glucose, and fructose) render them easily fermentable by a variety of intestinal bacteria in a non-selective way that results in the production of undesirable gases that can cause abdominal bloating and discomfort (Sánchez-Mata et al. 1998). To reduce the concentration of flatulence-causing oligosaccharides stachyose and raffinose in the skims, an α-galactosidase was applied to hydrolyze the glycosidic bonds within raffinose and stachyose (Figure 6.5).

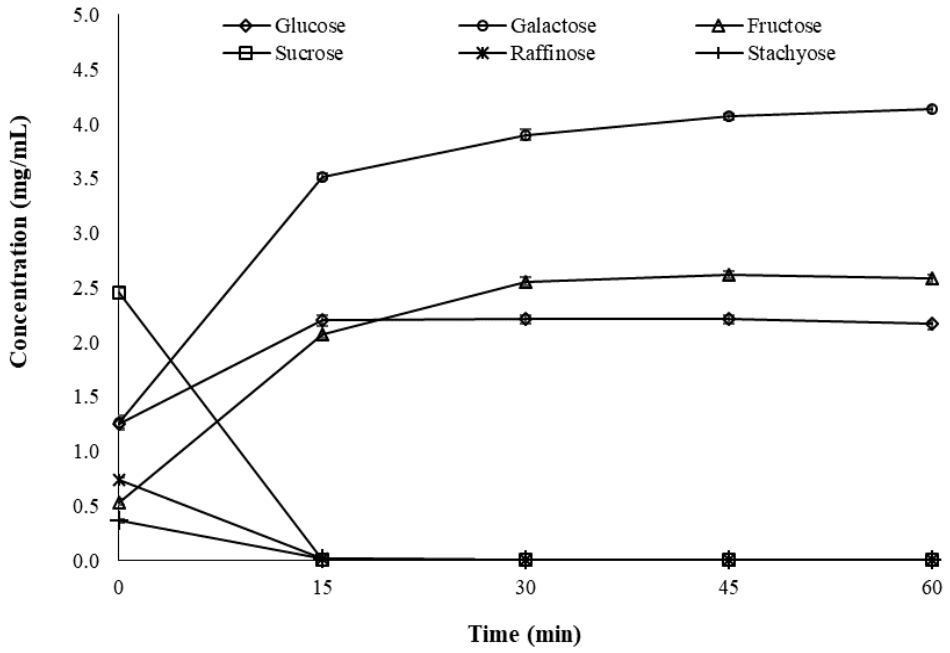


Figure 6.5. α-Galactosidase treatment on EAEP 1 skim. The carbohydrate concentrations were measured by HPAEC-PAD.

From Figure 6.5, we can observe that the α-galactosidase treatment completely hydrolyzed raffinose and stachyose in just 15 minutes, which was corroborated by the concurrent increase in the concentration of galactose released from the cleavage of the α-glycosidic bonds. Although sucrose concentration should increase with the α-galactosidase treatment, the observed decrease in sucrose reflects the hydrolysis of the glycosidic bonds between glucose and fructose by the α-galactosidase (De Moura et al. 2008), indicating that the enzyme preparation also possesses invertase activity. This can be observed by the simultaneous reduction in the sucrose concentration and increase in the glucose and fructose concentration. Our results are in agreement with the ones reported by De Moura Bell et al. 2013, who reported the complete reduction of stachyose in the protein extracts generated from AEP/EAEP of soybeans by the α-galactosidase treatment. Our

results demonstrate that the α-galactosidase treatment can certainly be an effective and fast treatment to reduce the presence of flatulence-causing oligosaccharides in the chickpea extracts and could therefore be introduced during the extraction process if adequate pH values are selected to favor the activity of the enzymes used.

6.3.4.3 Characterization of oligosaccharides in the skim fractions with LC-MS/MS

A total of 60 oligosaccharides were identified in the AEP and EAEP (1−3) skim fractions by inspecting the fragmentation patterns tandem MS spectra (Figure 6.6 and 6.7). With the masses of the precursor ions and fragment ions, the monosaccharide compositions of the oligosaccharides were determined. Among the 60 oligosaccharides, 46 contained only hexoses with a degree of polymerization range of 3 to 16 (Figure 6.6A–E). Stachyose, raffinose, and verbascose peaks, which were identified by comparing the retention times with the authentic standards, were the first three tallest peaks among the 46 hexose oligosaccharides (Figure 6.6A). As some oligosaccharides could originate from the enzyme formulations used in the EAEP treatments, the presence of oligosaccharides in the four enzymes used during the extraction was also examined. The results showed that no oligosaccharides were found in the alkaline protease, whereas 31 hexose oligosaccharides found in the skim fractions were also present in at least one of the three carbohydrases used in EAEP 2 and 3 (Figure 6.6C–E). It could therefore be confirmed that the remaining 12 oligosaccharides composed of 3–5 hexose residues and present in similar abundances in the AEP and the three EAEP skim fractions (Figure 6.6B–D) are endogenous oligosaccharides in chickpeas.

Ciceritol is a digalactosyl-pinitol present in the skim fractions in a high abundance (with peak areas close to stachyose; Figure 6.6F) firstly identified in chickpea (Quemener and Brillouet 1983). Besides ciceritol, seven oligosaccharides with relatively lower abundances possess similar

structures to ciceritol (Figure 6.6F). Among them, four are ciceritol isomers containing two hexose residues and one methyl-inositol (Hex_2+methyl-inositol); two are composed of three hexose residues and one methyl-inositol; one is composed of four hexose residues and one methyl-inositol. In addition to ciceritol, di-galactosyl-pinitol B and tri-galactosyl-pinitol A were identified in chickpeas in previous studies (L. Ruiz-Aceituno et al. 2017; Laura Ruiz-Aceituno et al. 2013). Although it was not possible to fully elucidate the exact structures (i.e., types of hexoses, methyl-inositols, and glycosidic linkages) of the other hexosyl-methyl-inositol derivatives, this work represents the first report of their existence in chickpeas, to the best of our knowledge.

Of interest, six oligosaccharides were exclusively found in the skim fractions generated by EAEP 2 and 3 (Figure 6.7). To ensure that the oligosaccharides aforementioned were generated from polysaccharide depolymerization under the action of the carbohydrase enzymes used during the extraction, we analyzed all the enzyme preparations used in the extraction. No oligosaccharides were identified in the enzyme preparations, confirming the de-novo origin of the said structures originated from the process. The monosaccharide composition of the six oligosaccharides, with signal intensities from high to low, were: Hex_4Pent_3, $Hex_3dHex_3Pent_3HexA_1$, Hex_2Pent_2, $Pent_7HexA_1$, Hex_5Pent_3, and $Hex_2dHex_3Pent_3Hex_1$. Because these oligosaccharides all contain multiple pentose units and other non-hexose monosaccharide units, they were not derived from the depolymerization of cellulose, which only consists of linear chains of β-1,4-linked glucan. Based on the monosaccharide compositions, it is likely that the oligosaccharides were generated from hemicellulose (e.g., xyloglucans) and pectin (e.g., from its component rhamnogalacturonan) partial hydrolysis (Tosh and Yada 2010; Wood et al. 2014; Yoo et al. 2012). Due to the lack of digestive enzymes in the human gastrointestinal tract able to break down glycosidic linkages of plant cell

wall polysaccharides, the newly generated oligosaccharides with diverse monosaccharide compositions in the skims of EAEP 2 and 3 could be novel prebiotics.

It was initially expected that more oligosaccharides could be generated through hydrolyzing polysaccharides by the carbohydrases used in EAEP 2 and 3, but in reality, only six oligosaccharides were found in the two skim samples in relatively low abundances. One plausible explanation for this result might be related to the polysaccharides' tangled structure and steric hindrance, which would in turn reduce enzyme accessibility. According to (Brummer et al. 2015), chickpea soluble fiber polysaccharides have a number average molecular weight (M_n) of 419 kDa and a weight average molecular weight (M_w) of 2,103 kDa. The massive size of the soluble polysaccharides may create steric hindrance issues for the endo-cleaving enzymes decreasing their accessibility and performance. Additionally, it is worth considering that some of the products generated by the enzymatic depolymerization of the polysaccharides might be larger than the size of oligosaccharides (3–20 monosaccharide units) and therefore not measured by LC-MS/MS. The reduction in molecular weight or the increase in solubility of cell wall polysaccharides could lead to a higher total carbohydrate content in the skims of EAEP 2 and 3 than EAEP 1, as the total carbohydrate quantification reported in Table 6.1. Moreover, the insoluble polysaccharides, which are more abundant than the soluble ones in chickpeas (Tosh and Yada 2010) might have even larger molecular sizes which are not possible to measure with the current analytical tools. Thus, a more intense enzymatic treatment might be needed to further hydrolyze those larger molecules into oligosaccharides. To further increase the concentration and diversity of novel oligosaccharides from chickpea polysaccharides, additional enzyme screening and process optimization would be needed.

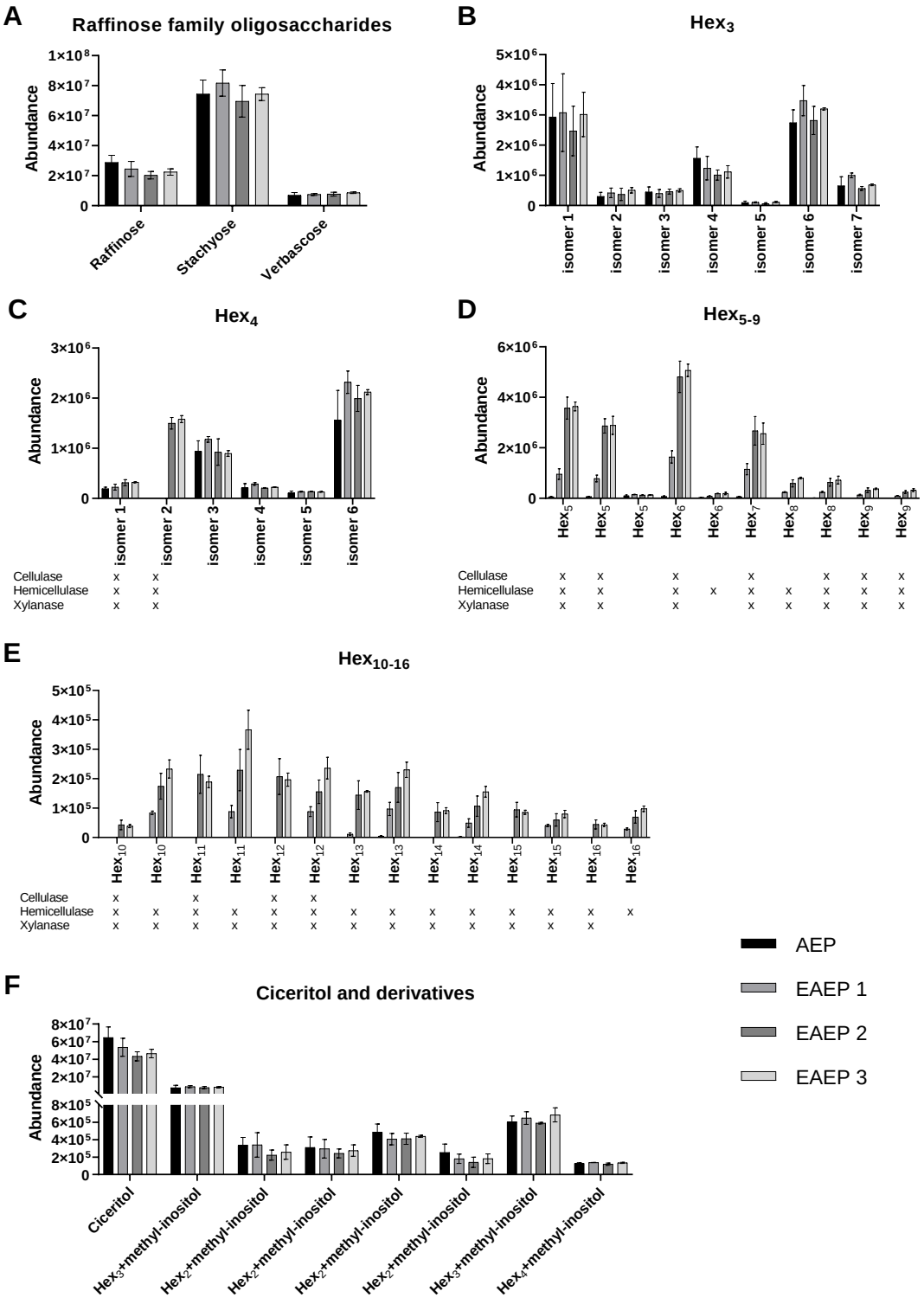


Figure 6.6. Relative quantification of oligosaccharides identified by LC-MS/MS in the skim fractions generated by AEP and EAEP (excluding the ones exclusively identified in EAEP 2 and EAEP 3).

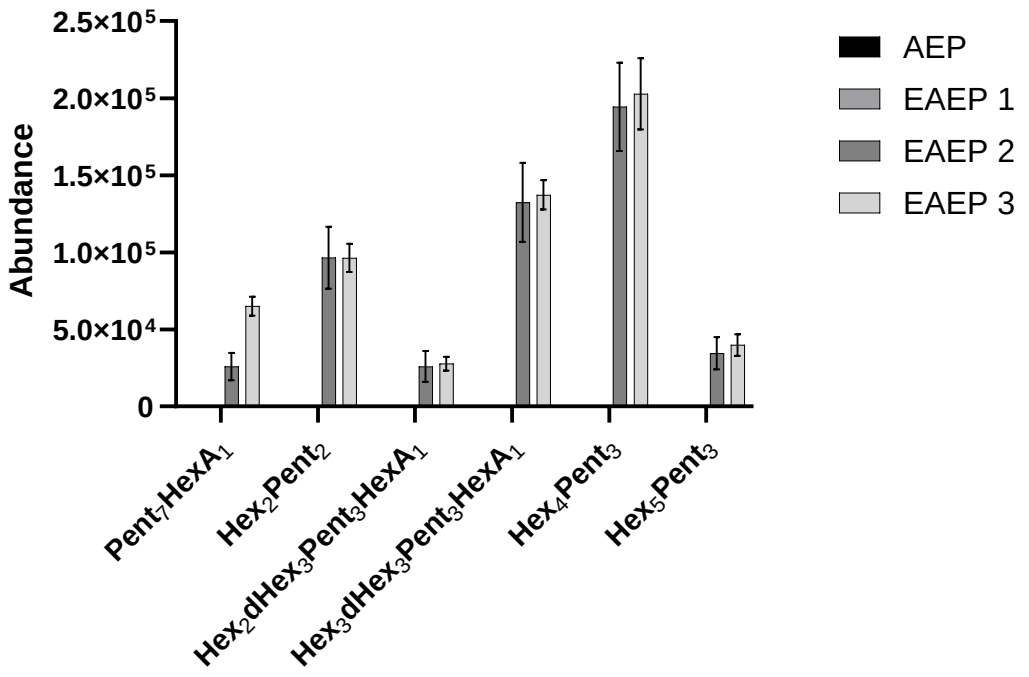


Figure 6.7. Relative quantification of oligosaccharides exclusively identified in skim fractions generated by EAEP 2 and EAEP 3.

6.3.5. Effects of enzymatic extraction on in vitro digestibility of skim proteins

Enzymatic extraction can significantly alter the *in vitro* digestibility of the extracted protein. In addition to being soluble, increased digestibility of chickpea proteins is another important functional property as it can promote nutritional benefits through higher intestinal absorption. The larger proteins observed in the AEP skim (section 6.3.2) can hinder digestibility, while the hydrolysis of these larger structures could improve its overall breakdown during human digestion (Sánchez-Vioque et al. 1999).

The *in vitro* digestibility of skim proteins from the non-enzymatic and the three enzymatic treatments are shown in Figure 6.8. The digestibility of all EAEP skim proteins was significantly

higher than that of the AEP skim (unhydrolyzed), highlighting the effectiveness of the use of enzyme during the extraction to enhance protein digestibility. Enzymatic hydrolysis significantly increased protein digestibility from 83.81% ± 1.86 (AEP), to 94.67% ± 8.70, 94.67% ± 6.47, and 90.79% ± 7.21 for the EAEP 1, 2, and 3, respectively, with no significant differences within the enzymatic treatments. As expected, the addition of carbohydrases did not significantly alter the *in vitro* digestibility of the EAEP skim proteins. Increased protein digestibility of EAEP skim proteins can be attributed to the breakdown of large proteins into smaller sizes by the protease (Souza et al., 2020; He et al., 2015), which corroborates with the DH and molecular weight results (Section 6.3.2). Clemente et al. (1998) reported that the digestibility of raw chickpeas increased from 71.8 ± 1.0 to 83.5 ± 0.1% after cooking, similar to the findings of Attia et al. 1994. Cooking chickpeas lead to protein denaturation and unfurling of the protein bodies, which improves the access to proteolysis by the saliva, gastric, and intestinal fluids. The digestibility of the AEP skim is similar to that of cooked chickpeas found by Clemente et al. 1998, which could be attributed to the steaming of the chickpeas before milling. Goertzen et al. 2020 reported digestibility values of 73.71 and 82.22% for untreated chickpea flour and isolate, respectively. However, no improvements in protein digestibility were observed when pepsin, trypsin, or papain were used to hydrolyze the chickpea protein isolate. It is worth mentioning that in their study, enzymes were used to hydrolyze the chickpea protein isolate while in our work, enzymes were used to extract proteins from the chickpea flour, which lead to a significant increase in protein digestibility (from 83.8 to 94.6%). Our results highlight that the use of selected enzymes to assist the extraction of full-fat chickpea flour is an effective strategy not only to improve protein extractability but to significantly enhance protein *in vitro* digestibility and solubility.

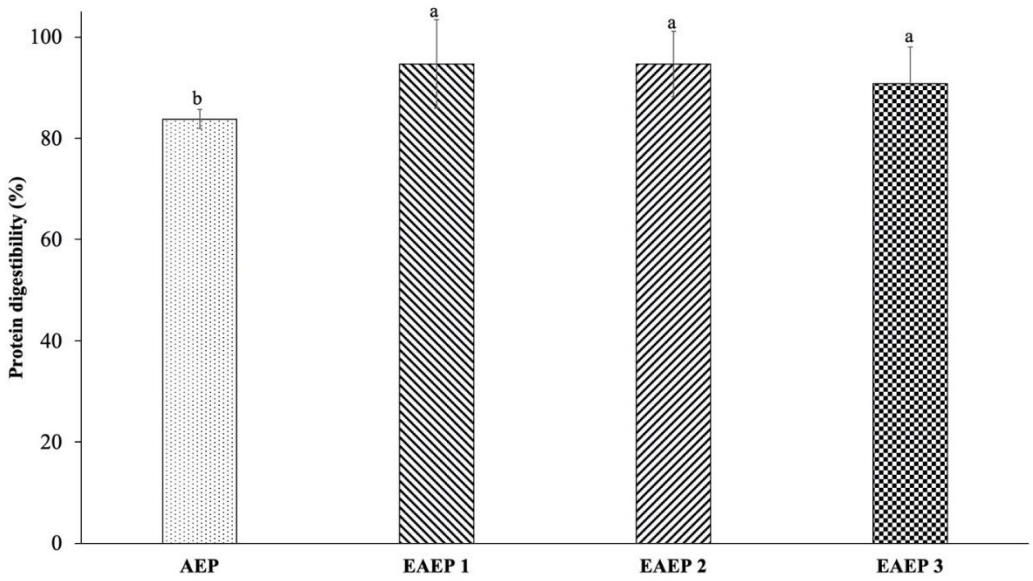


Figure 6.8. Effects of extraction processes (AEP and EAEP) on the in vitro digestibility of skim proteins. Different letters indicate statistically significant difference by one-way ANOVA followed by Tukey's test at $p<0.05$.

6.4. Conclusions

The extraction methods proposed in this work improved the overall protein extractability from full-fat chickpea flour and significantly enhanced the nutritional quality and functionality of the extracted proteins, without the need of performing upstream lipid removal by solvent extraction or mechanical pressing. The use of alkaline protease in the extraction increased oil (49.8 to 77.2%) and protein extractability (62.8 to 84.0%) from chickpea flour while releasing smaller and more soluble proteins. Proteolysis resulted in increased protein solubility at acidic pH (73%), where chickpea protein solubility is unfavored. Importantly, it did increase *in vitro* protein digestibility to 94.6%, which can therefore enhance the nutritional value of the extracted protein. While the use of carbohydrase pretreatments did not increase oil and protein extractability, it did release new

oligosaccharides as revealed by LC-MS/MS, warranting future investigation of the potential prebiotic properties of the novel oligosaccharides. The α-galactosidase post-extraction treatment eliminated the presence of flatulence-causing oligosaccharides in the extracts. These results demonstrate that enzymatic modifications can be exploited to provide the food industry with plant-based proteins that are highly functional, applicable, and produced by an environmentally friendly process. The impact of proteolysis on the sensory properties of food products containing chickpea hydrolyzates would merit further investigation.

Author's contribution

Kazunori Machida: Methodology, Data curation, Formal analysis, Writing - review & editing. **Yu-Ping Huang**: Methodology, Formal analysis, Data curation; Writing - review & editing. **Fernanda Furlan Goncalves Dias:** Methodology, Data curation, Writing - review & editing. **Daniela Barile:** Conceptualization, Data curation, Supervision, Funding, Writing - review & editing. **Juliana Maria Leite Nobrega de Moura Bell:** Conceptualization, Data curation, Supervision, Funding, Writing - review & editing.

Conflict of interest statement

The authors declare no conflict of interest.

Data availability statement

Data can be made available upon reasonable request.

Funding

This work was supported by the USDA Agriculture Research Service, Pulse Crop Health Initiative (PCHI 2019), Agreement number 58-3060-9-046; and USDA National Institute of Food and Agriculture, Hatch/Multi-State project [1023517].

References

Almeida, N. M. de, Dias, F. F. G., Rodrigues, M. I., & Bell, J. M. L. N. de M. (2019). Effects of Processing Conditions on the Simultaneous Extraction and Distribution of Oil and Protein from Almond Flour. *Processes 2019, Vol. 7, Page 844*, *7*(11), 844. https://doi.org/10.3390/PR7110844

Attia, R. S., Shehata, A. M. E. T., Aman, M. E., & Hamza, M. A. (1994). Effect of cooking and decortication on the physical properties, the chemical composition and the nutritive value of chickpea (Cicer arietinum L.). *Food Chemistry*, *50*(2), 125–131. https://doi.org/10.1016/0308-8146(94)90108-2

Bornhorst, G. M., & Singh, R. P. (2013). Kinetics of in Vitro Bread Bolus Digestion with Varying Oral and Gastric Digestion Parameters. *Food Biophysics*, *8*(1), 50–59. https://doi.org/10.1007/s11483-013-9283-6

Boulter, D., & Croy, R. R. D. (1997). The Structure and Biosynthesis of Legume Seed Storage Proteins: A Biological Solution to the Storage of Nitrogen in Seeds. *Advances in Botanical Research*, *27*(C), 1–92. https://doi.org/10.1016/S0065-2296(08)60280-3

Boye, J. I., Aksay, S., Roufik, S., Ribéreau, S., Mondor, M., Farnworth, E., & Rajamohamed, S. H. (2010). Comparison of the functional properties of pea, chickpea and lentil protein concentrates processed using ultrafiltration and isoelectric precipitation techniques. *Food Research International*, *43*(2), 537–546. https://doi.org/10.1016/j.foodres.2009.07.021

Brummer, Y., Kaviani, M., & 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

Campbell, K. A., & 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

Campbell, K. A., Glatz, C. E., Johnson, L. A., Jung, S., De Moura, J. M. N., Kapchie, V., & Murphy, P. (2011). Advances in aqueous extraction processing of soybeans. *JAOCS, Journal of the American Oil Chemists' Society*, *88*(4), 449–465. https://doi.org/10.1007/s11746-010-1724-5

Carbonaro, M., Cappelloni, M., Nicoli, S., Lucarini, M., & Carnovale, E. (1997). *Solubility-Digestibility Relationship of Legume Proteins*.

Chang, Y. W., Alli, I., Molina, A. T., Konishi, Y., & Boye, J. I. (2012). Isolation and Characterization of Chickpea (Cicer arietinum L.) Seed Protein Fractions. *Food and Bioprocess Technology*, *5*(2), 618–625. https://doi.org/10.1007/s11947-009-0303-y

Cheng, M. H., Sekhon, J. J. K., Rosentrater, K. A., Wang, T., Jung, S., & Johnson, L. A. (2018). Environmental impact assessment of soybean oil production: Extruding-expelling process, hexane extraction and aqueous extraction. *Food and Bioproducts Processing*, *108*, 58–68. https://doi.org/10.1016/j.fbp.2018.01.001

Clemente, A., Sánchez-Vioque, R., Vioque, J., Bautista, J., & Millán, F. (1998). Effect of cooking on protein quality of chickpea (Cicer arietinum) seeds. *Food Chemistry*, *62*(1), 1–6. https://doi.org/10.1016/S0308-8146(97)00180-5

De Moura Bell, J., Maurer, D., Yao, L., Wang, T., Jung, S., & Johnson, L. A. (2013). Characteristics of oil and skim in enzyme-assisted aqueous extraction of soybeans. *JAOCS, Journal of the American Oil Chemists' Society*, *90*(7), 1079–1088. https://doi.org/10.1007/s11746-013-2248-6

De Moura, J., & Johnson, L. A. (2009). Two-stage countercurrent enzyme-assisted aqueous extraction processing of oil and protein from soybeans. *JAOCS, Journal of the American Oil Chemists' Society*, *86*(3), 283–289. https://doi.org/10.1007/s11746-008-1341-8

De Moura, J. M.L.N., Campbell, K., Mahfuz, A., Jung, S., Glatz, C. E., & Johnson, L. (2008). Enzyme-assisted aqueous extraction of oil and protein from soybeans and cream de-emulsification. *JAOCS, Journal of the American Oil Chemists' Society*, *85*(10), 985–995. https://doi.org/10.1007/s11746-008-1282-2

De Moura, J. M.L.N., Maurer, D., Jung, S., & Johnson, L. A. (2011). Pilot-plant proof-of-concept for integrated, countercurrent, two-stage, enzyme-assisted aqueous extraction of soybeans. *JAOCS, Journal of the American Oil Chemists' Society*, *88*(10), 1649–1658. https://doi.org/10.1007/s11746-011-1831-y

De Moura, J. M L N, Campbell, K., De Almeida, N. M., Glatz, C. E., & Johnson, L. A. (2011). Protein recovery in aqueous extraction processing of soybeans using isoelectric precipitation and nanofiltration. *JAOCS, Journal of the American Oil Chemists' Society*, *88*(9), 1447–1454. https://doi.org/10.1007/s11746-011-1803-2

de Moura, Juliana Maria Leite Nobrega, de Almeida, N. M., & 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

de Souza, T. S. P., Dias, F. F. G., Oliveira, J. P. S., de Moura Bell, J. M. L. N., & Koblitz, M. G. B. (2020). Biological properties of almond proteins produced by aqueous and enzyme-assisted aqueous extraction processes from almond cake. *Scientific Reports*, *10*(1), 1–12. https://doi.org/10.1038/s41598-020-67682-3

Dias, F. F. G., Almeida, N. M., Souza, T. S. P., Taha, A. Y., & Bell, J. M. L. N. M. (2020). Characterization and Demulsification of the Oil-Rich Emulsion from the Aqueous Extraction Process of Almond Flour. *Processes*, *8*(10), 1228. https://doi.org/10.3390/pr8101228

Dias, F. F. G., & Bell, J. M. L. N. de M. (2022). Understanding the impact of enzyme-assisted aqueous extraction on the structural, physicochemical, and functional properties of protein extracts from full-fat almond flour. *Food Hydrocolloids*, *127*, 107534. https://doi.org/10.1016/J.FOODHYD.2022.107534

Duranti, M., & Gius, C. (1997). Legume seeds: Protein content and nutritional value. *Field Crops Research*, *53*(1–3), 31–45. https://doi.org/10.1016/S0378-4290(97)00021-X

Ghribi, A. M., Maklouf Gafsi, I., Sila, A., Blecker, C., Danthine, S., Attia, H., et al. (2015). Effects of enzymatic hydrolysis on conformational and functional properties of chickpea protein isolate. *Food Chemistry*, *187*, 322–330. https://doi.org/10.1016/j.foodchem.2015.04.109

Ghribi, A. M., Sila, A., Przybylski, R., Nedjar-Arroume, N., Makhlouf, I., Blecker, C., et al. (2015). Purification and identification of novel antioxidant peptides from enzymatic hydrolysate of chickpea (Cicer arietinum L.) protein concentrate. *Journal of Functional Foods*, *12*, 516–525. https://doi.org/10.1016/j.jff.2014.12.011

Goertzen, A. D., House, J. D., Nickerson, M. T., & Tanaka, T. (2020). The impact of enzymatic hydrolysis using three enzymes on the nutritional properties of a chickpea protein isolate. https://doi.org/10.1002/cche.10361

González, A. D., Frostell, B., & Carlsson-Kanyama, A. (2011). Protein efficiency per unit energy and per unit greenhouse gas emissions: Potential contribution of diet choices to climate change mitigation. *Food Policy*, *36*(5), 562–570. https://doi.org/10.1016/j.foodpol.2011.07.003

He, S., Simpson, B. K., Ngadi, M. O., & Ma, Y. (2015). In vitro studies of the digestibility of lectin from black turtle bean (Phaseolus vulgaris). *Food Chemistry*, *173*, 397–404. https://doi.org/10.1016/j.foodchem.2014.10.045

Huang, Y., Robinson, R. C., Furlan Goncalves Dias, F., Maria Leite Nobrega de Moura Bell, J., & 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 2022, Vol. 11, Page 340*, *11*(3), 340. https://doi.org/10.3390/FOODS11030340

Huang, Y.P., Robinson, R. C., & Barile, D. (2022). Food Glycomics: Dealing with Unexpected Degradation of Oligosaccharides during Sample Preparation and Analysis. *J. Food Drug Anal.*, 2022, 30 (1), 62–76. https://doi.org/10.38212/2224-6614.3393

Kim, T. K., Yong, H. I., Kim, Y. B., Jung, S., Kim, H. W., & Choi, Y. S. (2021). Effects of organic solvent on functional properties of defatted proteins extracted from Protaetia brevitarsis larvae. *Food Chemistry*, *336*. https://doi.org/10.1016/j.foodchem.2020.127679

Kou, X., Gao, J., Zhang, Z., Wang, H., & Wang, X. (2013). Purification and identification of antioxidant peptides from chickpea (Cicer arietinum L.) albumin hydrolysates. *LWT - Food Science and Technology*, *50*(2), 591–598. https://doi.org/10.1016/j.lwt.2012.08.002

L'hocine, L., Boye, J. I., & Arcand, Y. (2006). Composition and Functional Properties of Soy Protein Isolates Prepared Using Alternative Defatting and Extraction Procedures. *Journal of Food Science*, *71*(3), C137–C145. https://doi.org/10.1111/j.1365-2621.2006.tb15609.x

Masuko, T., Minami, A., Iwasaki, N., Majima, T., Nishimura, S. I., & Lee, Y. C. (2005). Carbohydrate analysis by a phenol-sulfuric acid method in microplate format. *Analytical Biochemistry*, *339*(1), 69–72. https://doi.org/10.1016/j.ab.2004.12.001

Nadar, S. S., Pawar, R. G., & Rathod, V. K. (2017). Recent advances in enzyme extraction strategies: A comprehensive review. *International Journal of Biological Macromolecules*, *101*, 931–957. https://doi.org/10.1016/J.IJBIOMAC.2017.03.055

Nielsen, P. M. (2001). Improved Method for Determining Protein Hydrolysis. *Journal of Food Science*, *66*(5), 642–646.

Quemener, B., & Brillouet, J. M. (1983). Ciceritol, a pinitol digalactoside form seeds of chickpea, lentil and white lupin. *Phytochemistry*, *22*(8), 1745–1751. https://doi.org/10.1016/S0031-9422(00)80263-0

Reese, E. T., Siu, R. G., & Levinson, H. S. (1950). The biological degradation of soluble cellulose derivatives and its relationship to the mechanism of cellulose hydrolysis. *Journal of bacteriology*, *59*(4), 485–497. https://doi.org/10.1128/jb.59.4.485-497.1950

Romero, J., Sun, S.-M. M., Mcleester, R. C., Bliss, F. A., & Hall, T. C. (1975). *Heritable Variation in a Polypeptide Subunit of the Major Storage Protein of the Bean, Phaseolus vulgaris L.'*. *Plant Physiol* (Vol. 56).

Ruiz-Aceituno, L., Carrero-Carralero, C., Ruiz-Matute, A. I., Ramos, L., Sanz, M. L., & Martínez-Castro, I. (2017). Characterization of cyclitol glycosides by gas chromatography coupled to mass spectrometry. *Journal of Chromatography A*, *1484*, 58–64. https://doi.org/10.1016/j.chroma.2017.01.001

Ruiz-Aceituno, Laura, Rodríguez-Sánchez, S., Ruiz-Matute, A. I., Ramos, L., Soria, A. C., & Sanz, M. L. (2013). Optimisation of a biotechnological procedure for selective fractionation of bioactive inositols in edible legume extracts. *Journal of the Science of Food and Agriculture*, *93*(11), 2797–2803. https://doi.org/10.1002/jsfa.6103

Sánchez-Mata, M. C., Peñuela-Teruel, M. J., Cámara-Hurtado, M., Díez-Marqués, C., & Torija-Isasa, M. E. (1998). Determination of Mono-, Di-, and Oligosaccharides in Legumes by High-Performance Liquid Chromatography Using an Amino-Bonded Silica Column. *Journal of Agricultural and Food Chemistry*. American Chemical Society. https://doi.org/10.1021/jf980127w

Sánchez-Vioque, R., Clemente, A., Vioque, J., Bautista, J., & Millán, F. (1999). Protein isolates from chickpea (Cicer arietinum L.): Chemical composition, functional properties and protein characterization. *Food Chemistry*, *64*(2), 237–243. https://doi.org/10.1016/S0308-8146(98)00133-2

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

Shaabani, S., Yarmand, M. S., Kiani, H., & Emam-Djomeh, Z. (2018). The effect of chickpea protein isolate in combination with transglutaminase and xanthan on the physical and rheological characteristics of gluten free muffins and batter based on millet flour. *LWT - Food Science and Technology*, *90*, 362–372. https://doi.org/10.1016/j.lwt.2017.12.023

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

Souza, T. S. P., Dias, F. F. G., Koblitz, M. G. B., & Bell, J. M. L. N. M. (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

Tosh, S. M., & Yada, S. (2010). Dietary fibres in pulse seeds and fractions: Characterization, functional attributes, and applications. *Food Research International*. https://doi.org/10.1016/j.foodres.2009.09.005

Tzitzikas, E. N., Vincken, J. P., De Groot, J., Gruppen, H., & Visser, R. G. F. (2006). Genetic variation in pea seed globulin composition. *Journal of Agricultural and Food Chemistry*, *54*(2), 425–433. https://doi.org/10.1021/jf0519008

U.S. Department of Agriculture, A. R. S. (2019). FoodData Central. *FoodData Central*.

Wallace, T. C., Murray, R., & Zelman, K. M. (2016). The Nutritional Value and Health Benefits of Chickpeas and Hummus. https://doi.org/10.3390/nu8120766

WHO. (2003). Diet, nutrition and the prevention of chronic diseases. *World Health Organization technical report series*, *916*.

Wood, J. A., Knights, E. J., Campbell, G. M., & Choct, M. (2014). Differences between easy- and difficult-to-mill chickpea (Cicer arietinum L.) genotypes. Part I: Broad chemical composition. *Journal of the Science of Food and Agriculture*, *94*(7), 1437–1445. https://doi.org/10.1002/jsfa.6437

Yoo, H. D., Kim, D., Paek, S. H., & Oh, S. E. (2012). Plant cell wall polysaccharides as potential resources for the development of novel prebiotics. *Biomolecules and Therapeutics*. Biomol Ther (Seoul). https://doi.org/10.4062/biomolther.2012.20.4.371

Yust, M., Pedroche, J., Millán-Linares, M. del C., Alcaide-Hidalgo, J. M., & Millán, F. (2010). Improvement of functional properties of chickpea proteins by hydrolysis with immobilised Alcalase. *Food Chemistry*, *122*(4), 1212–1217. https://doi.org/10.1016/j.foodchem.2010.03.121

Zayas, J. F. (1997). Chapter 1: Solubility of Proteins. In *Functionality of Proteins in Foods* (pp. 6–75). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-59116-7

Zhong, C., Wang, R., Zhou, Z., Jia, S. R., Tan, Z. L., & Han, P. P. (2012). Functional properties of protein isolates from caragana korshinskii Kom. Extracted by three different methods. *Journal of Agricultural and Food Chemistry*, *60*(41), 10337–10342. https://doi.org/10.1021/jf303442u

Chapter VII

Don’t throw away the cooking water: Aquafaba from chickpeas and common beans contains potentially bioactive oligosaccharides and peptides

Abstract

The cooking water of pulses, known as aquafaba, has recently become a sought-after food ingredient thanks to its foaming and emulsifying properties. Because pulse components might leach into the cooking water, we set out to identify bioactive compounds such as peptides and oligosaccharides, which can potentially modulate the composition and function of the gut microbiome and improve human health. Aquafaba was collected by draining canned and home-cooked chickpeas and common beans. Oligosaccharides and peptides were extracted via protein precipitation followed by solid-phase extraction, with different materials for oligosaccharides and peptides. Dimethyl labeling of peptides was conducted to facilitate small peptide analysis. With this technique, it was found that α- and γ-glutamyl peptides could be easily differentiated by the significant a_1 and b_1 fragment ions. A total of 433 and 350 peptides with varied abundance were identified in chickpea and common bean aquafaba, respectively. About 50 small α-peptides in chickpeas and common beans were found to be bioactive according to the bioactive peptide database BIOPEP. Many kokumi and anti-inflammatory peptides, including γ-Glu-Phe and γ-Glu-Tyr in chickpeas and γ-Glu-S-methyl-Cys and γ-Glu-Leu in common beans, were found in high abundance. Several novel γ-glutamyl peptide sequences were also identified. In terms of oligosaccharides, we identified 71 and 57 oligosaccharides in aquafaba from chickpeas and common beans, respectively. Stachyose, raffinose, and verbascose were major oligosaccharides in chickpeas and common beans, with ciceritol predominant uniquely in chickpeas. Oligosaccharides composed of 3−7 hexoses or 2−4 hexoses plus a methyl-inositol (only in chickpeas), pentose, acetyl-hexose, or phosphohexose were also found but in lower abundance. Considering that pulse processing generates massive amounts of aquafaba, the discovery of such a high number of bioactive peptides and oligosaccharides can promote its valorization and so improve the

sustainability of food systems while diverting biomolecules from low-cost streams into value-added ingredients that improve human health.

Keywords: LC-MS/MS; cyclitol-containing oligosaccharides; gamma-glutamyl peptides; dimethyl labeling; *Cicer arietinum*; *Phaseolus vulgaris*

7.1. Introduction

Legumes are nowadays considered valuable crops for improving human health while attaining more sustainable agricultural and food systems. Legumes provide various nutrients, such as proteins, dietary fiber, and minerals. Studies also demonstrated that consumption of legumes was associated with a lower risk of cardiovascular diseases (Afshin et al., 2014; Becerra-Tomás et al., 2019). Because of their ability to fix atmospheric nitrogen, legume cultivation requires a lower amount of nitrogen fertilizers, resulting in a lower greenhouse gas emission than other crops (Stagnari et al., 2017). Legumes are also an important alternative protein source to substitute animal-based protein foods for addressing the increasing global protein demand and reducing food production's environmental impact (Ismail et al., 2020).

Pulses are a subgroup of leguminous crops that are harvested as dry grains and used for human food and feed purposes, excluding those mainly used for oil extraction (e.g., soybeans and peanuts) (Food and Agriculture Organization of the United Nations (FAO), 2022). Many pulses, such as chickpeas (*Cicer arietinum*) and common beans (*Phaseolus vulgaris*), are often soaked and then cooked in water before consumption. This process helps soften the texture, increase digestibility, and reduce the activity of the antinutritional factors which are typically present. Canning (also known as retort processing) is the conventional method used by the food industry

at the industrial scale, and it generally consists of various steps, including soaking, blanching, can filling, thermal processing, and cooling (Schoeninger et al., 2017; USA Pulses, 2022; W. Wang et al., 2021). The processing conditions, such as temperature, time length, and pressure, according to the properties of the pulse grains and the equipment used, are optimized to attain good product quality and achieve commercial sterility (Schoeninger et al., 2017; USA Pulses, 2022; W. Wang et al., 2021). During the cooking and canning process, some pulse components naturally leach out into the cooking water. Although the cooking water is often considered waste and disposed of, novel usages are being sought to leverage its excellent functional properties. The cooking water of legumes is known as *aquafaba*, which is derived from the Latin words *aqua* meaning water, and *faba* meaning beans. The functional foaming properties of aquafaba were first discovered in chickpea aquafaba by Joël Roessel, a vegan French musician (He et al., 2021). Besides the foaming properties, the gelling and emulsification properties of aquafaba from chickpeas and other legumes have been studied and reported in the scientific literature (He et al., 2021). These functional properties are believed to be assocaited to several components in aquafaba, such as protein, soluble polysaccharides, oligosaccharides, saponins, and phenolic compounds (Mustafa & Reaney, 2020). Due to these properties, aquafaba has been proposed as an ingredient to be used as an egg replacer for producing vegan desserts (He et al., 2021; Mustafa et al., 2018).

Besides being utilized for its functional properties in gastronomical applications, discovering healthful components in aquafaba could be another route for upcycling the cooking water for additional implementations. Oligosaccharides and bioactive peptides are known to exist in a variety of foods. Non-digestible oligosaccharides have been gaining interest because they can stimulate the growth of beneficial microorganisms in the human gut, which are believed to play a critical role in modulating human health conditions. Peptides with specific structures have been

shown to exhibit various physiological functions that benefit human health, from lowering blood pressure to acting as natural antimicrobials. This study aimed to characterize oligosaccharides and peptides in aquafaba from chickpeas and common beans to identify the potentially bioactive constituents.

7.2. Materials and methods

7.2.1. Materials

Trifluoroacetic acid (TFA), formaldehyde (37%), sodium cyanoborohydride, 25% ammonia solution (LC-MS grade) were purchased from MilliporeSigma (St. Louis, MO, USA). Acetonitrile (LC-MS grade) and formic acid (LC-MS grade) were purchased from Fisher Scientific (Waltham, MA, USA). Sodium cyanoborohydride was obtained from TCI (Tokyo, Japan). Peptide standards were obtained from Bachem (Torrance, CA, USA).

Seven aquafaba samples (four chickpeas and three common beans) were collected by draining canned chickpeas or common beans purchased from a grocery store in Davis, CA, USA; one chickpea aquafaba sample was collected after cooking chickpeas using a home pressure cooker (6 qt, InstantPot). Before cooking, 150 g dried chickpeas were soaked in 300 mL tap water at room temperature for 12 h. The soaked chickpeas, the soaking water, and an additional 200 mL of water were added to the pressure cooker and cooked at high pressure for 15 min, after which the cooking water was collected for further analyses.

7.2.2. Sample preparation

7.2.2.1. Protein precipitation

Fifty microliters of aquafaba were mixed with 100 μL cold ethanol. After incubating at −30 °C for 1 h, the mixture was centrifuged at 13,000 g, 4 °C for 30 min. The supernatant was

transferred to new tubes and dried in a centrifugal evaporator (MiVac Quattro, Genevac Ltd., Ipswitch, Suffolk, UK) at 30 °C. The dried deproteinized samples were redissolved in 50 μL of water.

7.2.2.2. Solid-phase extraction

7.2.2.2.1. Oligosaccharide purification

Oligosaccharides were purified with two consecutive SPE steps (reverse-phase/strong cation exchange (RP/SCX) and porous graphic carbon (PGC)) for eliminating peptides and desalting, respectively, as previously described (Huang, Robinson, Dias, et al., 2022). Briefly, RP/SCX SPE cartridges (Strata-X-C, 30 mg, 1 mL tube volume, Phenomenex, Torrance, CA, USA) were activated with 2 mL acetonitrile and equilibrated with 2 mL of 0.1% formic acid in water (v/v). Ten microliters of deproteinized aquafaba samples dissolved in water were mixed with 40 μL of 0.1% formic acid in water and loaded onto the conditioned cartridges. The cartridges were flushed with 1.5 mL of 0.1% formic acid in water for collecting the eluates containing oligosaccharides. The eluates were subsequently loaded to a PGC SPE microplate with 40 μL chromatographic media beds (Glygen Corporation, Columbia, MD, USA) preconditioned with 300 μL of 80% acetonitrile, 0.1% TFA in water (v/v/v) and 300 μL of water. The microplate wells were then washed with 1.2 mL of water. Oligosaccharides were then eluted by flushing the microplate with 600 μL of 40% acetonitrile in water (v/v) and then 600 μL of 40% acetonitrile, 0.1% TFA in water (v/v/v) for collecting neutral and charged oligosaccharides, respectively. The samples were then dried in a centrifugal evaporator.

7.2.2.2.2. Peptide purification

Peptide purification was carried out as previously described (Huang, Dias, et al., 2022) with slight modification. Briefly, C18 solid-phase extraction cartridges (Discovery DSC-18, 500

mg, 3 mL tube volume, Sigma) were activated by 5 mL of acetonitrile and equilibrated by 5 mL of 0.1% TFA (v/v). After being acidified by adding TFA, samples were loaded onto the cartridges, which were then washed with 6 mL of 0.1% TFA. The peptides were eluted by 4 mL 80% acetonitrile, 0.1% TFA in water (v/v/v). The collected eluates were dried in a centrifugal evaporator.

7.2.2.3. Dimethyl labeling

The dried supernatant samples (collected after protein precipitation) were dissolved in 100 μL of 100 mM sodium acetate buffer (pH 5.5). Ten microliters of 4% formaldehyde and 10 μL of 600 mM $NaBH_3CN$ were added to the sample solution, vortexed, and incubated at room temperature for 2 h. After acidifying the samples by adding 1 μL of TFA, the samples were purified by solid-phase extraction as described in section 7.2.2.2.2.

7.2.3. Oligosaccharide quantification

Quantification of oligosaccharides was conducted on a Dionex ICS-5000+ high-performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD) system (ThermoFisher, Sunnyvale, CA, USA) equipped with a CarboPac PA200 guard column (3 × 50 mm) and a CarboPac PA200 analytical column (3 × 250 mm) (ThermoFisher). The deproteinized samples were appropriately diluted and injected into the HPAEC-PAD system. Chromatographic separation was carried out with a 15-min isocratic elution using 50 mM NaOH at a flow rate of 0.5 mL/min. The column was flushed with 200 mM NaOH for 5 min at the end of each run and equilibrated with 50% NaOH for 10 min before the next injection. The concentration of sucrose, raffinose, stachyose, and verbascose in samples was calculated with calibration curves constructed with external standards (0.1−10 μg/mL for raffinose and verbascose and 0.1−20 μg/mL for sucrose and stachyose).

7.2.4. Liquid chromatography-quadrupole-time-of-flight analysis

LC-MS/MS analysis for oligosaccharide and peptide identification was performed on an Agilent 6520 Accurate-Mass Q-TOF LC-MS with a Chip Cube interface (Agilent Technologies, Santa Clara, CA).

7.2.4.1. Oligosaccharide identification

The mobile phase was composed of 5 mM ammonium acetate, 3% acetonitrile in water (solvent A) and 5 mM ammonium acetate, 90% acetonitrile in water (solvent B). Using ammonium acetate as a mobile phase additive can promote the formation of ammonium species, which can aid the differentiation of authentic oligosaccharides and in-source fragments and consequently avoid incorrect identification (Huang, Robinson, & Barile, 2022). The oligosaccharide samples were delivered to the enrichment column of an Agilent PGC-Chip II (G4240-64010) with 100% A at a flow rate of 4 μL/min. The oligosaccharides were separated on the analytical column of the PGC chip with a 60-min gradient. The gradient started from 100% A, increased from 0 to 16 % B in 20 min, from 16to 44% B in 10 min, from 44 to 100% B in 5 min, and was held at 100% B for 10 min. The system was equilibrated at 100% A for 15 min before 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 1875 V. The ions were scanned within the range of *m/z* 150–2500 at a rate of 1 spectrum sec^{-1}. The four most abundant ions in each MS analysis cycle were isolated for tandem MS analysis with ramped collision energy (CE; $CE = 0.02 \times m/z - 3.5$). The active exclusion was enabled. Throughout the analysis, reference ions *m/z* 922.009798 and *m/z* 1221.990637 were used for continual mass calibration.

7.2.4.2. Peptide identification

Medium-sized peptides were characterized by analyzing nonderivatized peptide samples; small peptides were identified by analyzing both nonderivatized and dimethyl labeling samples (Huang, Dias, et al., 2022). To avoid an excessively high capillary pump pressure caused by injecting the aquafaba peptide samples onto a high-capacity chip, we firstly carried out the medium-sized peptides analysis with an Agilent Zorbax 300SB-C18 capillary chip (40 nL enrichment column; 75 μm ×150 mm, 5 μm analytical column). The peptide samples were then passed through a centrifugal filter with an MWCO of 3 kDa (Amicon) at 14,000 g, 4 °C for 30 min to remove components that potentially caused the high capillary pump pressure. The filtrates were used for small peptide analysis by using a high-capacity Agilent Polaris-HR-Chip (360 nL enrichment column; 75 μm × 150 mm analytical column), which has superior retention of small peptides. The mobile phase consisted of 3% ACN with 0.1% formic acid (v/v) (A) and 89.9% ACN with 0.1% formic acid (v/v) (B). The samples were delivered to the enrichment column of the Zorbax chip and the Polaris-HR-Chip with 100% A at flow rates of 4 and 2.5 μL min^{-1}, respectively. Peptide separation on both chips was performed at a flow rate of 0.3 μL min^{-1} with a 65 min-gradient programmed as follows: 0–30% B from 0–40 min; 30–45% B from 40–45 min; 45–100% B from 45–45.1 min; 100% B from 45.1–50 min; 100–0% B from 50–50.01 min; 0% B from 50.01–65 min. The mass spectrometer was operated by using the settings described by (Huang, Dias, et al., 2022) with capillary voltages of 1860 and 1940–1945 for the Zorbax chip and the Polaris-HR-Chip, respectively. Reference ions *m/z* 322.048121and *m/z* 922.009798 were used for continual mass calibration throughout the analysis.

7.2.5. Data analysis

7.2.5.1. Oligosaccharide identification

Oligosaccharides and their monosaccharide composition were identified by manually inspecting the fragmentation pattern in MS/MS spectra. To avoid identifying in-source fragments as authentic oligosaccharides, identification of oligosaccharides was only confirmed when the MS/MS spectra were generated from ammonium adduct precursor ions or from protonated precursor ions that have evident corresponding ammonium ions in the ESI mass spectra.

7.2.5.2. Peptide identification

Data analysis for peptide identification was performed on PEAKS Studio X Pro (Bioinformatics Solutions Inc., Waterloo, ON, Canada), followed by manual inspection. Peptides comprising 2–5 and ≥5 amino acid residues were analyzed with *de novo* sequencing and database search, respectively. For both *de novo* sequencing and database search, the mass error tolerance for precursor and fragment ions were 15.0 ppm and 0.02 Da, respectively, and the enzyme setting was none.

In the *de novo* sequencing, variable methylation (+14.0157) on Cys was allowed for both nonderivatized and dimethyl labeled samples; modifications related to dimethyl labeling (fixed modification of dimethylation (+28.0313) on N-terminal residues except for Lys and Pro, double dimethylation (+56.0626) on N-terminal Lys, and methylation (+14.0157) on N-terminal Pro; variable modification of dimethylation (+28.0313) on Lys at any position) was applied to dimethyl labeled samples only. The tandem MS spectra of the *de novo* peptide sequences with scores above 30 were manually inspected to determine the correct identification, with full-length sequencing, among the top 10 candidates. During the manual inspection, peptide sequences containing Lys at any position except for N-termini were only accepted when the Lys were dimethylated.

Database searches were conducted using the UniProt database (including both Swiss-Prot and TrEMBL) of organisms of *Cicer arietinum* and *Phaseolus vulgaris*, for chickpea and common bean samples, respectively, with no enzyme and unspecific digestion mode. Variable modification allowed included oxidation (+15.99) on Met, deamidation (+0.98) on Asn and Gln, and phosphorylation (+79.97) on Ser, Thr, and Tyr. After filtering at a false positive rate of 1%, peptide sequences with low-quality matching spectra and with parent proteins not in the top protein list were manually removed.

7.2.6. Bioactivity annotation

The identified peptide sequences were searched against the BIOPEP-UWM bioactive peptide database (http://www.uwm.edu.pl/biochemia/index.php/en/biopep, accessed 2021/1/16) to annotate their bioactivities that had been previously reported in the literature (Minkiewicz et al., 2019). Because BIOPEP-UWM primarily comprises α-peptides, a literature search was conducted to find potential activities for γ-glutamyl peptides. For peptides found by *de novo* sequencing containing undistinguishable isomeric Leu/Ile, the activities of any possible sequences were reported.

7.3. Results and discussion

7.3.1. Oligosaccharides

7.3.1.1. Quantification of major oligosaccharides by HPAEC-PAD

Fig. 7.1 shows the HPAEC-PAD chromatograms of the low-molecular-weight carbohydrates in aquafaba from chickpeas and common beans. Sucrose and the major oligosaccharides, including raffinose, stachyose, and verbascose, were identified by comparing the retention times of the authentic standards with the peaks present in the samples and then quantified using standard curves. The concentrations of raffinose, stachyose, and verbascose are presented in

Fig. 7.2. Stachyose content was the highest among the three oligosaccharides, ranging from 3.5 ± 0.1–6.3 ± 0.3 mg/mL and 4.6 ± 0.2–6.2 ± 0.4 mg/mL in chickpea and common bean aquafaba, respectively. Raffinose content in chickpea aquafaba (1.1 ± 0.1–1.7 ± 0.0 mg/mL) was slightly higher than in common bean aquafaba (0.44 ± 0.02–0.87 ± 0.01 mg/mL). Verbascose content was the lowest among the three oligosaccharides being quantified for all the chickpea (0.15 ± 0.01–0.24 ± 0.02 mg/mL) and most of the common bean (0.27 ± 0.06–0.60 ± 0.00 mg/mL) aquafaba samples. The relative abundance of the three oligosaccharides was in line with their concentrations in the whole legumes reported in previous studies (Díaz-Batalla et al., 2006; Elango et al., 2022; Siva et al., 2020; N. Wang et al., 2010). The concentration of sucrose, a simple sugar lacking prebiotic activity, was also quantified for comparison (Figure 1). Sucrose content was similar in concentration to the total raffinose family oligosaccharides (RFO; i.e., raffinose, stachyose, and verbascose) for both chickpea (5.9 ± 0.3–7.2 ± 0.2 mg/mL of sucrose and 6.1 ± 0.3–8.2 ± 0.2 mg/mL of total RFO) and common bean (4.5 ± 0.3–5.4 ± 0.3 mg/mL of sucrose and 5.6 ± 0.3–7.3 ± 0.5 mg/mL of total RFO) aquafaba. A large variation in stachyose content was observed among chickpea aquafaba from different brands of canned chickpeas, possibly due to the differences in the raw materials and in the processing procedures. The chickpea aquafaba sample prepared with a home pressure cooker presented sucrose and oligosaccharides levels comparable to canned chickpeas. Others (N. Wang et al., 2010) have reported that the process of cooking significantly reduced sucrose, raffinose, stachyose, and verbascose levels in chickpeas and common beans, in agreement with the results presented herein, which confirmed that a significant portion of oligosaccharides is indeed being released into the cooking water.

An oligosaccharide comprising two galactose residues and one pinitol has been previously reported to exist in chickpeas and was named ciceritol, from the scientific name of chickpeas,

Cicer arietinum (Quemener & Brillouet, 1983; Xiaoli et al., 2008). However, in the current study, the peak of ciceritol could not be identified solely by comparing the retention time due to the lack of commercially available standards. In (Xiaoli et al., 2008)'s study, sucrose and α-galactooligosaccharides in chickpeas were analyzed by high-performance chromatography with a refractive index detector (HPLC-RI) equipped with an aminopropyl column; with that analytical set-up, the peak of ciceritol eluted between raffinose and stachyose peaks. Nonetheless, in the current study's HPAEC-PAD chromatograms of chickpea aquafaba, which were generated with the CarboPac PA200 column, no other prominent peak was observed near raffinose and stachyose peaks (Fig. 7.1B), which was similar to the results reported in another study using HPAEC-PAD to analyze chickpea sugars and oligosaccharides (Gangola et al., 2014). According to (Gangola et al., 2014), two cyclitols, including *myo*-inositol and galactinol (1-α -D-galactosyl-*myo*-inositol), rapidly eluted within 2 min from the start of the gradient (20–34.4 mM NaOH from 0–2 min), which is considerably sooner compared with the appearance of monosaccharides, when analyzed with a CarboPac PA100 column. (Borges et al., 2006) also revealed that several cyclitol compounds eluted much earlier than monosaccharides when separated on a CarboPac PA1 column. Because ciceritol also consists of a cyclitol (pinitol) in the structure, it was suggested that ciceritol also had weak retention on a CarboPac PA200 column that was used in the current study. According to the LC-MS/MS analysis results, ciceritol was found only in the aquafaba from chickpeas, not in common beans (discussed in the next section). When comparing the HPAEC-PAD chromatograms of chickpea and common bean aquafaba (Fig. 7.1B and C, respectively), a prominent peak eluted at 2.1 min (labeled with an asterisk in the figure) was exclusively found in the chickpea sample. Based on the early elution characteristic demonstrated in prior studies as

described above and the differences observed between our chickpea and common bean chromatograms, the peak at 2.1 min was tentatively identified as ciceritol.

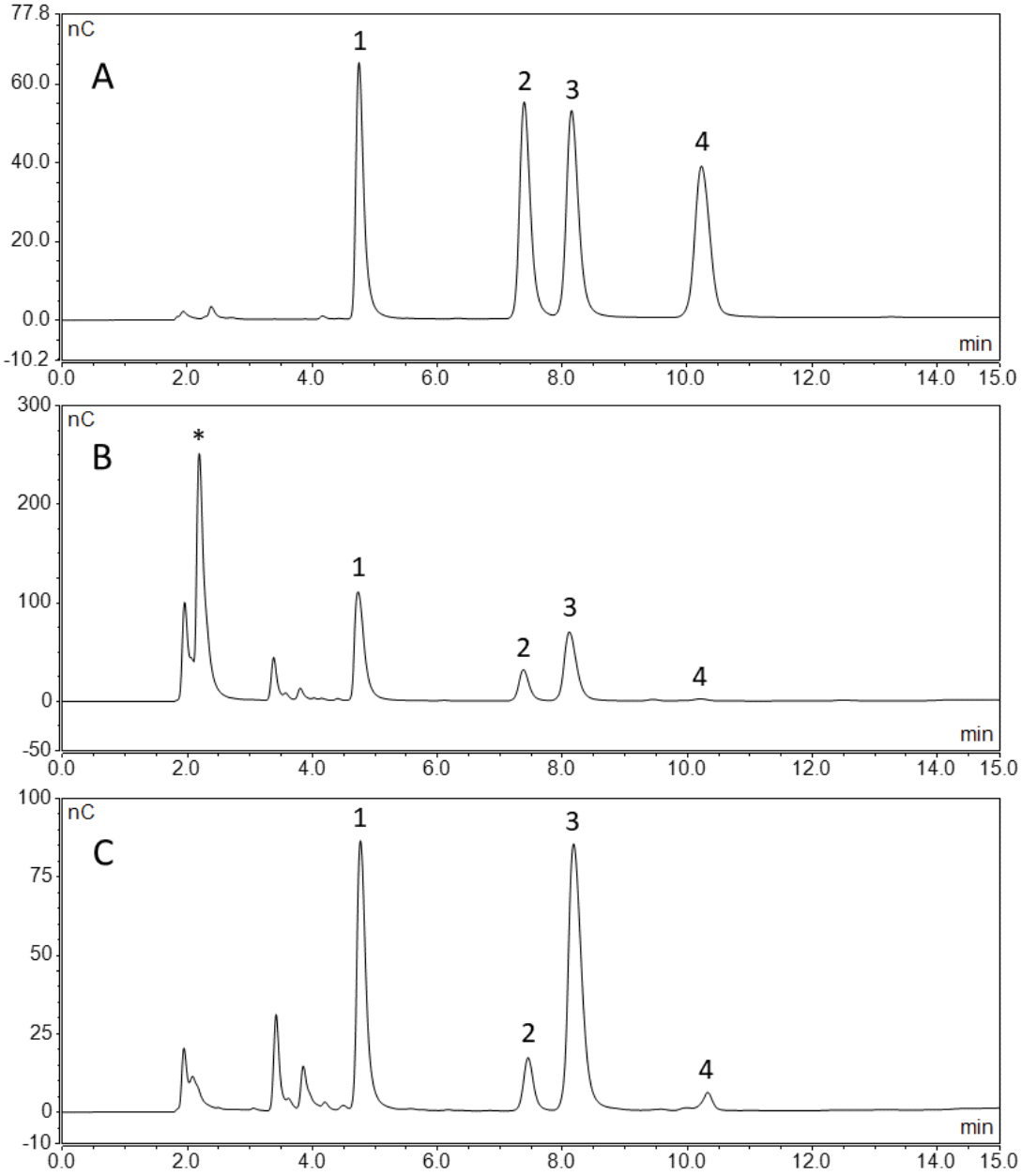


Fig. 7.1. HPAEC-PAD chromatograms of a standard mixture of sucrose and raffinose oligosaccharides (A) and aquafaba from chickpeas (B, data of the sample prepared with a pressure cooker) and common beans

(C, data of the white bean sample) separated on a CarboPac PA200 column. The y-axis represents signal intensity in nanocoulombs (nC); the x-axis represents retention time in minutes. Peak identities: 1, sucrose; 2, raffinose; 3, stachyose; 4, verbascose. The peak eluting at 2.1 minutes and labeled with an asterisk was tentatively identified as ciceritol.

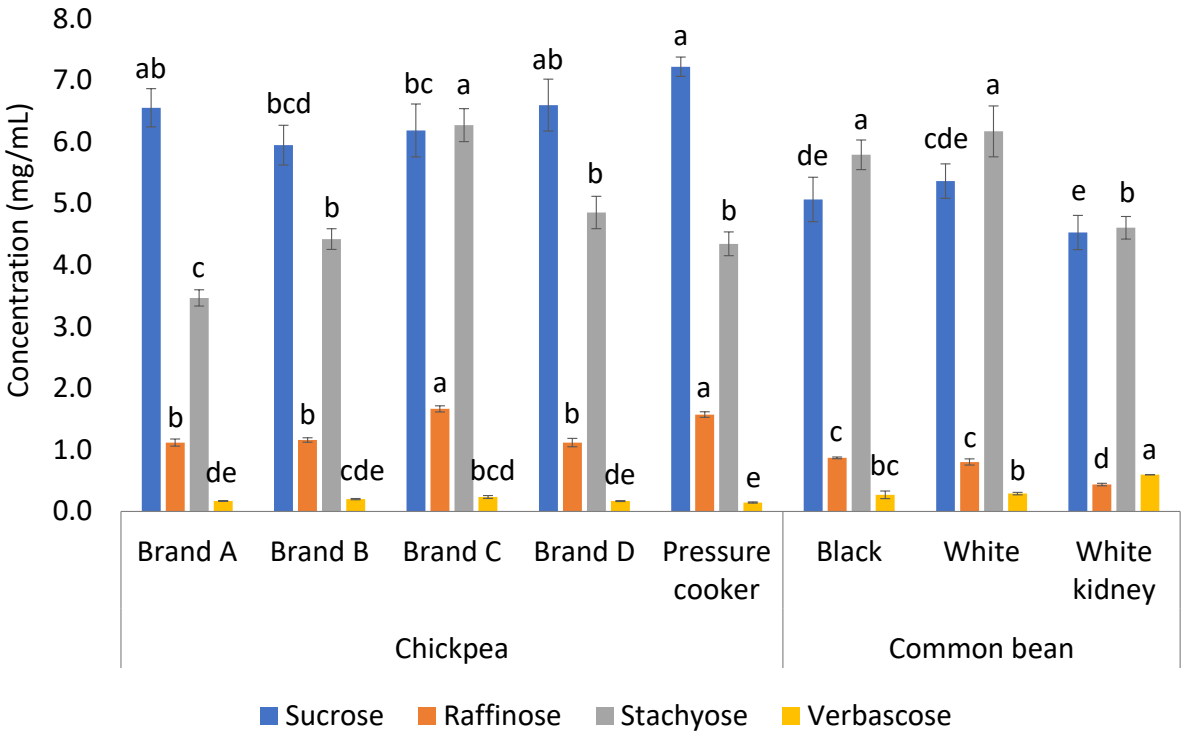


Fig. 7.2. Quantification of sucrose and raffinose-family oligosaccharide in chickpea and common bean aquafaba by HPAEC-PAD analysis. Values are means ± standard deviation (n = 3). Different lowercase letters indicate a significant difference ($p < 0.05$) of the same molecule among different samples by one-way analysis of variance (ANOVA) followed by Tukey's test.

7.3.1.2. Oligosaccharide profiling with LC-MS/MS

Besides quantifying the major oligosaccharides for which commercial standards exist, we also explored the presence of undiscovered oligosaccharides. Using optimized methods in LC-MS/MS mode, 74 and 63 oligosaccharide structures were identified in aquafaba generated from chickpeas and common beans, respectively (Table 7.1 and Table 7.S1). In one of the aquafaba samples from common beans (white beans), an additional 41 oligosaccharides with the monosaccharide composition $Hex_{4–17}$ were exclusively identified. Because the canned white bean sample included an extra plant ingredient, "granulated onion," the hexose oligosaccharides uniquely identified in this sample might actually originate from onion, which is known to contain fructooligosaccharides (Oku et al., 2019). Therefore, these oligosaccharides were not considered part of the common bean aquafaba oligosaccharides and were not included in Table 7.1 and 7.S1. Most chickpea and common bean aquafaba oligosaccharides were identified in the neutral fraction; in the charged oligosaccharide fraction, ten phosphoryl oligosaccharides containing three to four hexose residues were identified. A variety of oligosaccharide structures were measured in the neutral fraction, with a majority of the constituent monosaccharides being represented by hexoses, including $Hex_{3–7}$, $Hex_{2–4}Pent_1$, $Hex_{2–3}HexOAc_1$, and $Hex_{2–4}Glycerol_1$. The most abundant oligosaccharide in both the chickpea and common bean aquafaba was stachyose, followed by raffinose and verbascose (Fig. 7.3), which agrees with the trend that emerged from our HPAEC-PAD results. Notably, when comparing the monosaccharide compositions and retention times of the identified oligosaccharides, a considerable overlap of 58 oligosaccharides between the chickpea and common bean aquafaba was found.

The main difference between the two types of aquafaba was that oligosaccharides containing methyl-inositol were exclusively found in chickpea aquafaba (Table 7.1 and Fig. 7.3),

including three compounds with the monosaccharide composition of Hex_2Methyl-inositol$_1$, three Hex_3Methyl-inositol$_1$, and three Hex_4Methyl-inositol$_1$. Among the methyl-inositol-containing oligosaccharides identified in chickpea aquafaba, ciceritol (Gal_2Methyl-inositol$_1$), di-galactosyl-pinitol B (Gal_2Methyl-inositol$_1$), and tri-galactosyl pinitol A (Gal_3Methyl-inositol$_1$) have been identified in chickpeas in previous studies (Quemener & Brillouet, 1983; Ruiz-Aceituno et al., 2017). Our group previously characterized the oligosaccharides in chickpea flour extracts and identified five Hex_2Methyl-inositol$_1$, two Hex_3Methyl-inositol$_1$, and three Hex_4Methyl-inositol$_1$ (Machida et al., 2022). The different numbers of the methyl-inositol containing oligosaccharides with various monosaccharide compositions in our current and previous studies might be due to the variation between the chickpea samples used in the two studies (e.g., chickpea varieties) as well as the various mobile phase additives used in the LC-MS/MS analysis. The mobile phase used in our previous study (Machida et al., 2022) contained 0.1% formic acid, whereas, in the current study, 5 mM ammonium acetate was added to the eluent to help discriminate authentic oligosaccharides, which could be recognized with the formation of ammonium adduct ions, from in-source fragments (Huang, Robinson, & Barile, 2022). Therefore, although the different mobile phase additives might affect the chromatographic separation and lead to some variation in oligosaccharide identification, the results in the current study originate from the optimized LC-MS method and are considered more reliable thanks to the decreased chance of identifying in-source fragments as authentic oligosaccharides.

Table 7.1. Numbers of different types of oligosaccharides identified in chickpea and common bean aquafaba.

Oligosaccharides	Chickpea	Common bean
Neutral oligosaccharides		
$Hex_{3–7}$	42	39
$Hex_{2–4}Methyl\text{-}inositol_1$	9	0
$Hex_{2–4}Pent_1$	4	5
$Hex_{2–3}HexOAc_1$	6	6
$Hex_{2–4}Glycerol_1$	3	3
Charged oligosaccharides		
$Hex_{3–4}P_1$	10	10
Total	74	63

Hex: hexose; Pent: pentose; HexOAc: acetyl-hexose; P: phosphorylation.

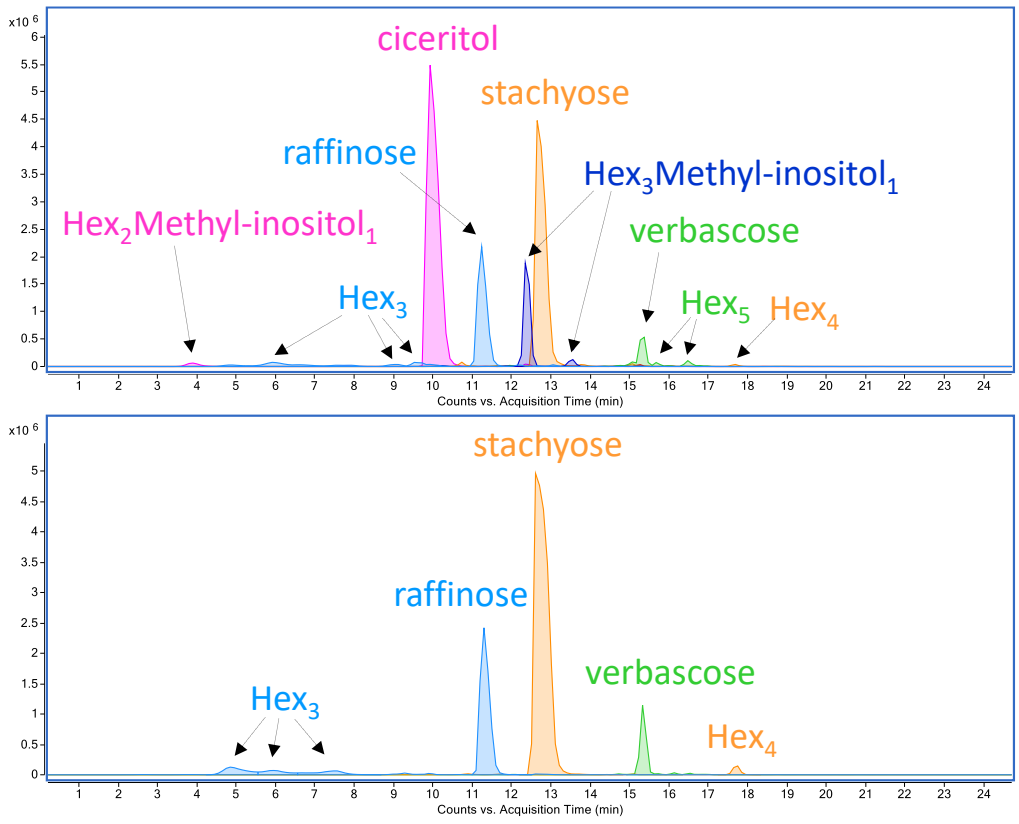


Fig. 7.3. LC-Q-TOF extracted ion chromatograms (EIC) of major oligosaccharides in aquafaba from chickpeas (upper panel, data of brand A) and common beans (bottom panel, data of black beans).

7.3.2. Peptides

7.3.2.1. Identification of medium-sized peptides

Food-derived peptide sequences containing at least five amino acid residues and up to 50 amino acids generally can be identified through LC-MS/MS analysis followed by database search. With this approach, 309 and 179 medium-sized peptide sequences were identified in chickpea and common bean aquafaba, respectively. The length of the identified medium-sized peptides ranged between 5–45 and 5–41 amino acid residues for the aquafaba from chickpeas and common beans, respectively. Medium-sized peptides found in chickpea aquafaba were mainly derived from vicilin-like protein (73 peptides), legumin J-like protein (32 peptides), and legumin A-like protein (26 peptides); for common bean aquafaba, the identified peptides were mainly derived from phaseolin (76 peptides) (Fig. 7.4). Legumin-like 11S globulin and vicilin-like 7S globulin are the major storage proteins in chickpeas with a ratio of about 3.6:1 (Tavano & Neves, 2008). Phaseolin (a vicilin-like 7S globulin) is the major seed storage protein of common beans (Lioi, 1989; Taylor et al., 2008). It is plausible that a certain level of proteolytic breakdown would occur in the raw legume seeds and during the cooking process, which would further release the generated peptides into the water during cooking/canning and storage. According to the BIOPEP-UWM database, none of the medium-sized peptide sequences were previously reported to be bioactive (Minkiewicz et al., 2019).

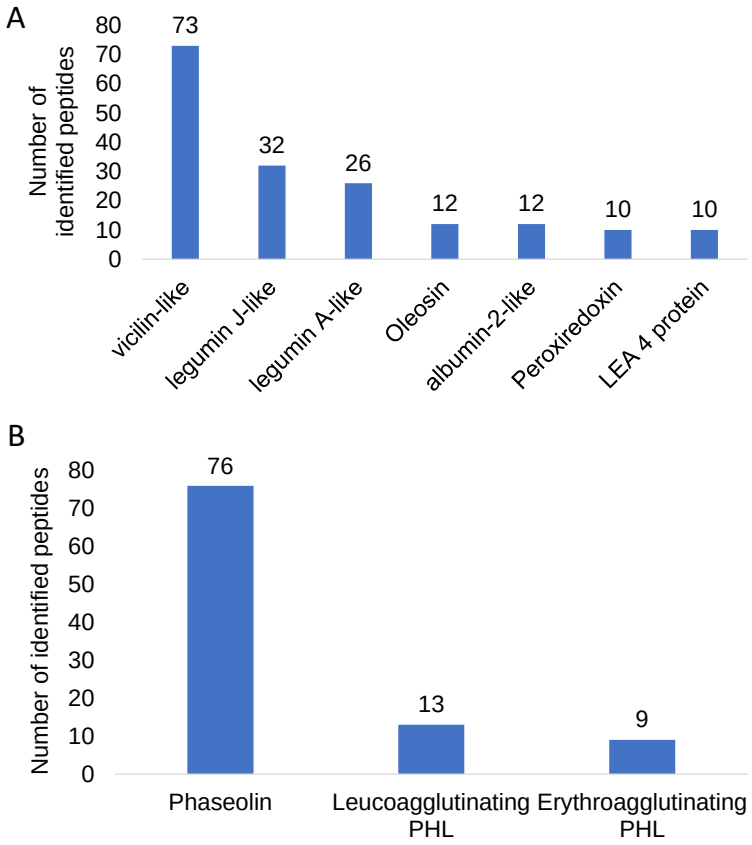


Fig. 7.4. Major parent proteins generating the medium-sized peptides identified in chickpea (A) and common bean (B) aquafaba. PHL: phytohemagglutinin.

7.3.2.2. Small peptides

7.3.2.2.1. Differentiation of α- and γ-glutamyl peptides

In a previous study (Huang, Dias, et al., 2022), our group showed that dimethyl labeling can be used as a simple and affordable technique to achieve signal enhancement of a_1 ions in tandem MS analysis and thus effectively aid in full-length sequencing of small peptides. In the current study, we observed that dimethyl labeled γ-glutamyl peptides resulted in unique fragmentation patterns which were distinct from α-peptides' patterns and thus enabled their

differentiation (Fig. 7.5 shows the chemical structures of γ- and α-glutamyl dipeptides). γ-Glutamyl peptides were found in legume seeds and some other foods (Lu et al., 2021; Clayton J. Morris et al., 1963; C J Morris & Thompson, 1962). Through LC-MS/MS analysis and the subsequent *de novo* sequencing, we identified several small peptides with N-terminal glutamine in aquafaba from chickpeas and common beans. Most of these glutamyl peptides produced an intense b_1 ion peak (*m/z* 158.081) when they were analyzed in dimethyl labeled forms (Fig. 7.6). For example, dimethylated dipeptides $(CH_3)_2$-EM, $(CH_3)_2$-EV, and $(CH_3)_2$-EL in common bean samples all had a prominent b_1 ion peak (*m/z* 158.081), with a much lower or even undetectable a_1 ion signal (*m/z* 130.086), in their tandem MS spectra (Fig. 7.6A–C). These patterns were evidently different from dimethylated α-dipeptides, which usually only contain a predominant a_1 ion peak (*m/z* 130.086) (Fig. 7.6D) (Huang, Dias, et al., 2022). In comparison, when analyzing nonderivatized peptides, signal intensities of a_1 and b_1 ions were much lower and, therefore, may not be used as the foundation to distinguish γ-glutamyl peptides from α-peptides (Fig. 7.6E–H). The identification of γ-glutamyl dipeptides (e.g., γ-EM, γ-EV, and γ-EL) was confirmed by comparing the fragmentation patterns of the corresponding peptide standards after dimethyl labeling. Therefore, it was evidenced that the different fragmentation patterns were useful in differentiating α- and γ-glutamyl peptides. In particular, significant a_1 and b_1 ions generated from dimethylated peptides could be used as diagnostic ions for confirming the identification of α- and γ-glutamyl peptides, respectively.

The distinct fragmentation patterns of dimethylated α- and γ-glutamyl peptides could be explained by hypothetical fragmentation pathways. Because methyl groups are electron donating, the electron density of N-terminal amines would increase after dimethyl labeling. In our analysis, small peptides with 2–5 amino acid residues usually formed singly charged ions after electrospray

ionization, except for those containing basic amino acid residues (i.e., Arg, Lys, and His). Based on the mobile proton model (Paizs & Suhai, 2005), in a dimethyl labeled small peptide ion, the added proton would be sequestered by the N-terminal amine due to the enhanced electron density. Also, the sequestered proton would be difficult to be mobilized from the N-terminal amine to other protonation sites, which are energetically and/or kinetically less favored. It was suggested that mobilizing the added proton from the N-terminal amine to the vicinal N-terminal amide bond would be less difficult than to the farther amide bonds. Therefore, the dissociation would mainly occur at the N-terminal amide bond. For α-peptides, the N-terminal amide bond dissociation is usually accompanied by CO neutral loss to form a_1 ions (Fig. 7.7A) (Hsu & Chen, 2016; Huang, Dias, et al., 2022). For γ-glutamyl peptides, it was hypothesized that the CO neutral loss and the consequent a_1 ion formation are less favored because of the farther distance between the N-terminal amines' lone pair electrons and the CO; therefore, the formation of b_1 ion is more favored (Fig. 7.7B).

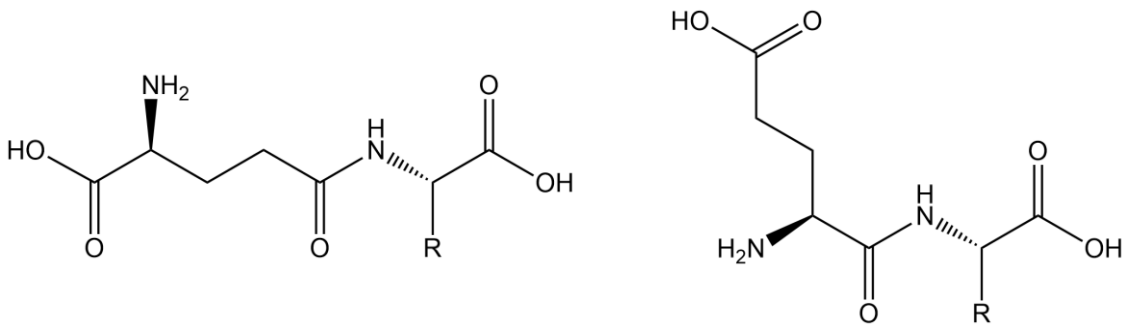


Fig. 7.5. Structures of γ-glutamyl (left) and α-glutamyl (right) dipeptides. R represents amino acid side chains.

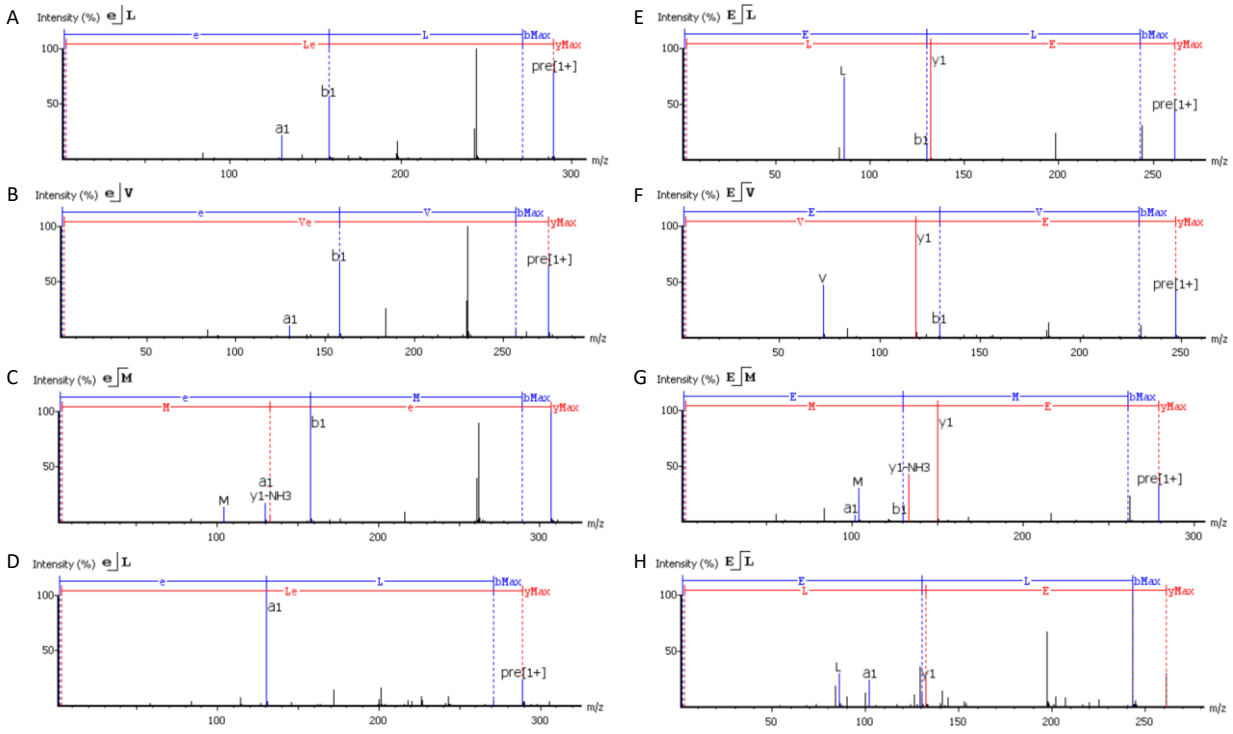


Fig. 7.6. Comparison of fragmentation patterns of example γ-glutamyl (A–C and E–G) and α-glutamyl (D and H) dipeptides in common bean aquafaba (data of black beans and white kidney beans) analyzed in dimethylated (A–D) and nonderivatized (E–H) forms. Dipeptides corresponding to the spectra: γ-Glu-Leu (A, E), γ-Glu-Val (B, F), γ-Glu-Met (C, G), and α-Glu-Lxx. Lxx represents Leu or Ile. Dimethylated γ-glutamyl peptides (A–C) generated more intense b_1 ions (*m/z* 158.081) than a_1 ions (*m/z* 130.086), whereas dimethylated α-glutamyl dipeptides (D) generated mainly a_1 ions (*m/z* 130.086).

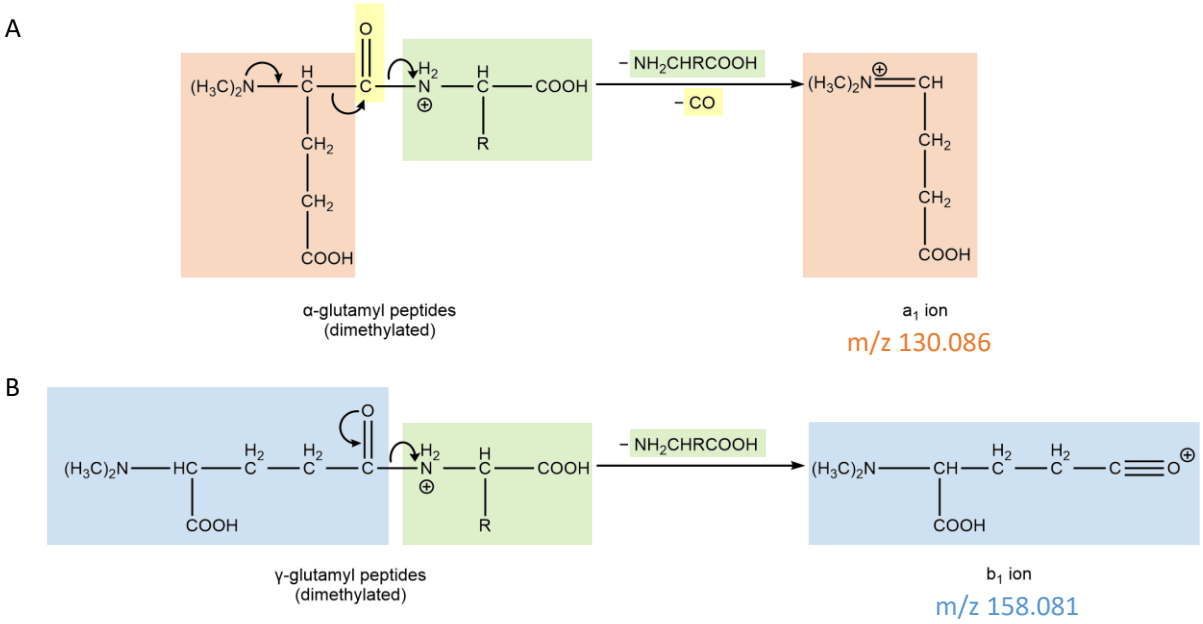


Fig. 7.7. Tendency of forming a_1 and b_1 ions during collision-induced dissociation (CID) fragmentation of dimethylated glutamyl peptides. The hypothesized fragmentation is exemplified by α- and γ-glutamyl dipeptides. Dimethylated α-glutamyl dipeptides tend to form a_1 ions (*m/z* 130.086) (A), whereas dimethylated γ-glutamyl dipeptides tend to form b_1 ions (*m/z* 158.081) (B). R represents amino acid side chains.

7.3.2.2.2. Identification of small peptides

Small peptides comprising 2–5 amino acid residues were identified through *de novo* sequencing in aquafaba from chickpeas (Table 7.2) and common beans (Table 7.3), in which 124 and 174 small peptides, respectively, including both α-peptides and γ-glutamyl peptides, were found. Among them, some sequences with a length of five amino acid residues (LSFDN in chickpeas and GLLGL, PFYFN, and PGFPN in common beans) were also identified via protein database search. By comparing extracted ion chromatograms (EIC) of the identified peptides, it

was found that γ-glutamyl peptides were present in a much higher abundance than α-peptides. The relatively low abundances of α-peptides might suggest there were low degrees of protein hydrolysis and/or low levels of α-peptides leach into the cooking water. EIC of major γ-glutamyl peptides (Fig. 7.8) demonstrated that their compositions were substantially different in chickpea and common bean aquafaba. Chickpea aquafaba was rich in γ-Glu-Tyr and γ-Glu-Phe, whereas common bean aquafaba consisted of abundant γ-Glu-S-methyl-Cys and γ-Glu-Leu. γ-Glutamyl peptides containing S-methyl-Cys were exclusively identified in common bean aquafaba and absent in chickpea aquafaba, possibly due to the existence of critical genes and enzymes involved in their biosynthesis in common beans (Liao et al., 2013; Saboori-Robat et al., 2019; Yang et al., 2019). γ-Glutamyl peptides naturally exist in legume seeds as small molecules and are not conjugated with proteins. Thus, similar to oligosaccharides, they could leach into the water during cooking. Previous studies reported the presence of some γ-glutamyl peptides in common bean seeds, including γ-Glu-S-methyl-Cys, γ-Glu-Leu, γ-Glu-Val, γ-Glu-Met, γ-Glu-Cys-β-Ala (homoglutathione), and γ-Glu-S-methyl-Cys-β-Ala (S-methylhomoglutathione) (Dunkel et al., 2007; Liao et al., 2013; Clayton J. Morris et al., 1963). γ-Glu-Tyr and γ-Glu-Phe were identified in chickpeas (Mekky et al., 2015). The peptides that had been previously identified were generally present in relatively high abundances in the aquafaba samples (Fig. 7.8). Besides them, several γ-glutamyl peptides were also found. To the best of our knowledge, they were identified in common beans and chickpeas for the first time. For example, γ-Glu-Glu-S-methyl-Cys, γ-Glu-S-methyl-Cys-Glu-S-methyl-Cys, γ-Glu-Glu-Lxx (Lxx represents Leu or Ile), γ-Glu-Lxx-Glu-S-methyl-Cys, and γ-Glu-S-methyl-Cys-Glu-Lxx, in common bean aquafaba and γ-Glu-Glu-Tyr and γ-Glu-Glu-Phe in chickpea aquafaba were confirmed to contain an N-terminal γ-Glu due to the significant b_1 ion peak in the tandem MS spectra (Fig. 7.9).

Table 7.2. Amino acid sequence and characteristics of the small peptides (2–5 amino acids) identified in chickpea aquafaba through *de novo* sequencing.

Peptide[1]	Dimethylated									Nonderivatized								
	Denovo score[2]	z[3]	RT (min)	a_1 or b_1	Brand A	Brand B	Brand C	Brand D	Brand E	Denovo score	z	RT (min)	a_1 or b_1	Brand A	Brand B	Brand C	Brand D	Brand E
AAGP	60	1	4.62		−	−	−	−	+					−	−	−	−	−
AEL	38	1	12.96		+	+	+	+	+	72	1	11.15		+	+	−	+	+
AEW					−	−	−	−	−	72	1	18.96		+	+	+	+	+
AFE	34	1	12.00		+	−	−	+	−					−	−	−	−	−
AL	41	1	10.14		−	−	+	−	+					−	−	−	−	−
ALE	38	1	8.96		−	−	+	−	−					−	−	−	−	−
ALNVN	65	1	14.53		+	−	−	+	−					−	−	−	−	−
ALP	35	1	16.75		−	−	+	−	−					−	−	−	−	−
ATL	40	1	12.73		−	−	−	−	+					−	−	−	−	−
AVL	35	1	17.25		−	−	−	−	+					−	−	−	−	−
DL	56	1	11.15		+	−	+	+	+					−	−	−	−	−
DLGVT	85	1	18.20		−	−	+	−	−					−	−	−	−	−
DNLLH	89	1	22.81		−	+	−	−	−					−	−	−	−	−
DVL	70	1	18.29		−	−	+	−	−					−	−	−	−	−
γ-EAW[4]	32	1	20.40	b_1	+	−	−	−	−					−	−	−	−	−
γ-ECA	34	1	5.18	b_1	−	+	−	−	−					−	−	−	−	−
γ-ECE	48	1	5.90	b_1	−	+	−	+	−	60	1	4.69	tiny b_1	−	−	−	+	−
γ-ECV	45	1	11.86	b_1	−	−	+	−	−					−	−	−	−	−

γ-EEEF	99	1	18.57	b_1	+	−	+	−	+					−	−	−	−	−
γ-EEF	77	1	17.65	b_1	+	+	+	+	+	79	1	16.36	b_1/a_1	+	+	+	+	+
γ-EEW	75	1	20.33	b_1	+	−	+	−	−	69	1	19.00	b_1/a_1	+	−	+	−	+
γ-EEY	76	1	11.20	b_1	+	+	+	+	+	68	1	10.03	b_1/a_1	+	−	+	−	+
γ-EF	49	1	15.67	b_1	−	+	+	−	+	69	1	13.58	b_1	+	+	+	+	+
γ-EFR	77	2	11.72	b_1	−	−	+	−	−	76	2	9.20	b_1	+	+	+	−	+
EFW					−	−	−	−	−	71	1	32.34	no a_1/b_1	−	+	+	−	−
α-EL	45	1	11.06	a_1	+	+	+	+	+	73	1	7.88	a_1	−	−	+	−	+
γ-EL	79	1	12.86	b_1	+	−	−	+	−					−	−	−	−	−
γ-EL	82	1	13.94	b_1	+	+	+	+	+	84	1	11.57	b_1	+	+	+	+	+
γ-EM	50	1	8.84	b_1	+	+	−	+	+					−	−	−	−	−
EQLLH					−	−	−	−	−	97	1	22.89	no a_1/b_1	+	+	−	−	−
γ-ESF	43	1	11.24	b_1	+	+	+	+	+					−	−	−	−	−
γ-EV	72	1	8.32	b_1	+	+	−	+	+	82	1	6.61	b_1/a_1	+	+	−	+	+
γ-EW	75	1	18.92	b_1	+	+	+	−	+	80	1	17.21	b_1	+	+	+	+	+
γ-EY	49	1	9.59	b_1	+	+	+	+	+	81	1	8.32	b_1	+	+	+	+	+
γ-EYR					−	−	−	−	−	67	2	5.46	b_1	+	−	−	+	−
FGLN	66	1	21.74		+	−	−	−	−					−	−	−	−	−
FL	41	1	24.76		+	+	+	+	+	73	1	20.97		+	−	+	+	+
FLD	40	1	18.45		+	+	+	+	+	71	1	17.09		+	+	+	+	+
FLDR	56	2	13.46		−	−	+	+	−	98	2	13.18		+	+	+	+	+
FLED	57	1	18.86		+	−	−	−	−					−	−	−	−	−

FLN	36	1	16.31	+	+	+	+	+	70	1	15.29	+	+	+	+	+
FP	31	1	17.93	−	−	+	−	−				−	−	−	−	−
FTNV	57	1	15.54	+	+	−	−	−	83	1	15.33	+	+	−	−	−
GF	46	1	11.75	+	+	−	+	+	58	1	9.27	−	−	+	−	+
GFEE	81	1	13.99	+	−	−	+	−				−	−	−	−	−
GFEQ	42	1	13.62	−	−	−	−	+				−	−	−	−	−
GL	54	1	9.14	+	+	+	+	+	61	1	7.35	−	−	+	+	+
GLD	38	1	6.79	−	−	+	+	−				−	−	−	−	−
GLK	71	2	4.73	+	−	−	+	−				−	−	−	−	−
GLL	40	1	21.68	−	−	+	−	−				−	−	−	−	−
GLLGL	75	1	32.47	−	−	−	−	+				−	−	−	−	−
GLP	46	1	15.27	−	−	+	−	+				−	−	−	−	−
GLTP	81	1	14.38	−	−	+	−	−				−	−	−	−	−
GSYT	89	1	7.13	+	−	−	−	−				−	−	−	−	−
GVF	32	1	19.80	−	−	−	−	+				−	−	−	−	−
GVGY	83	1	13.69	−	−	+	+	+				−	−	−	−	−
GVLTV	89	1	25.54	−	−	−	−	+				−	−	−	−	−
GY	35	1	6.90	−	−	+	−	−				−	−	−	−	−
GYLN	78	1	13.60	−	−	−	−	+				−	−	−	−	−
LD	37	1	5.76	−	+	+	+	+				−	−	−	−	−
LDL	46	1	24.91	−	−	+	−	+	74	1	22.30	−	−	+	−	−
LE	40	1	6.28	−	−	+	−	+	69	1	4.81	−	−	−	−	+
LED	37	1	8.07	−	−	+	−	−	73	1	6.16	−	−	+	−	−

LEL	36	1	25.32	−	−	−	−	+	65	1	21.16	−	−	−	−	+
LLEP	97	1	19.31	−	−	−	−	+	92	1	18.97	−	−	−	−	+
LLPLR	76	2	22.16	+	−	−	−	−	98	2	20.58	+	−	−	−	−
LP 1	45	1	10.33	−	−	+	−	+	69	1	9.47	−	−	−	−	+
LP 2				−	−	−	−	−	70	1	10.36	−	−	−	−	+
LPAVL				−	−	−	−	−	79	1	30.13	+	−	−	−	−
LRD				−	−	−	−	−	77	2	4.16	+	−	−	−	−
LSEVN	71	1	12.26	−	+	−	+	+				−	−	−	−	−
LSF	32	1	27.44	+	−	−	−	−				−	−	−	−	−
LSFD	67	1	24.82	+	+	+	+	+	97	1	21.99	+	+	−	+	+
LSFDD	75	1	24.67	−	+	−	−	+				−	−	−	−	−
LSFDN	72	1	22.56	+	+	+	+	+	97	1	19.89	+	+	+	+	+
LSLN	76	1	19.26	+	−	−	−	+	93	1	16.79	+	−	−	+	−
LV	35	1	10.55	−	−	+	−	−				−	−	−	−	−
LVH	69	2	4.61	−	−	−	−	+				−	−	−	−	−
MG	36	1	14.34	+	+	−	−	−				−	−	−	−	−
NL	75	1	9.69	−	−	+	−	−				−	−	−	−	−
NPGGP				−	−	−	−	−	96	1	5.82	−	−	+	−	−
PFSL	78	1	29.50	−	+	−	+	+	97	1	28.89	−	+	−	+	+
PGL				−	−	−	−	−	61	1	12.25	−	−	−	−	+
PGLTL	64	1	26.90	−	−	−	−	+				−	−	−	−	−
PLKY	97	2	14.75	−	−	−	−	+				−	−	−	−	−
PLR				−	−	−	−	−	67	2	4.56	−	−	−	−	+

PLY	35	1	17.18	+	−	−	−	−				−	−	−	−	−
PLYSN	61	1	12.94	+	−	−	−	−	95	1	12.53	−	−	−	−	+
PRSF				−	−	−	−	−	93	2	8.83	−	−	−	−	+
PVAAY	74	1	15.77	−	−	−	−	+				−	−	−	−	−
SAQL	52	1	12.49	+	−	−	−	+				−	−	−	−	−
SEVD	58	1	5.41	+	−	−	−	−				−	−	−	−	−
SEVN	82	1	4.44	−	+	+	+	+				−	−	−	−	−
SF	36	1	11.96	−	−	+	−	+				−	−	−	−	−
SFDN	72	1	9.39	−	+	−	−	+	94	1	7.46	−	−	−	−	+
SL	61	1	9.58	+	+	+	+	+	74	1	6.65	−	−	+	−	−
SLE	50	1	8.61	−	−	+	−	−				−	−	−	−	−
SLF	31	1	25.51	−	−	−	−	+				−	−	−	−	−
SLL	61	1	22.01	−	−	−	−	+				−	−	−	−	−
SLNLK				−	−	−	−	−	93	2	14.24	+	−	−	−	−
SLP	69	1	16.09	−	−	+	−	−				−	−	−	−	−
SLSL	92	1	24.54	+	+	+	−	+	97	1	22.72	−	−	−	−	+
SV	35	1	5.63	−	−	−	+	−				−	−	−	−	−
SVE	42	1	5.22	−	−	−	−	+				−	−	−	−	−
SVELN	86	1	15.28	+	−	−	−	+				−	−	−	−	−
TF	35	1	15.57	−	−	+	−	−				−	−	−	−	−
TL 1	52	1	11.90	−	−	+	−	+				−	−	−	−	−
TL 2	70	1	12.78	−	−	+	−	+	80	1	7.08	−	−	+	−	+
TLF				−	−	−	−	−	73	1	23.2	−	−	−	−	+

TLVNA	89	1	16.27	+	+	−	+	+	92	1	14.89	−	−	−	+	−
TT	36	1	3.67	−	−	+	−	+				−	−	−	−	−
TV	78	1	7.69	−	−	+	−	−				−	−	−	−	−
TVLL	50	1	28.34	+	−	−	−	−				−	−	−	−	−
TVP	33	1	14.21	−	−	+	−	−				−	−	−	−	−
VAAP	78	1	12.41	+	+	−	+	+				−	−	−	−	−
VAVL	60	1	24.18	−	−	−	−	+	95	1	21.98	−	−	−	−	+
VD	38	1	3.97	−	−	+	−	−				−	−	−	−	−
VF	33	1	16.27	−	−	−	−	+				−	−	−	−	−
VGL				−	−	−	−	−	56	1	16.44	−	−	−	−	+
VL	41	1	12.20	−	−	+	−	−	74	1	9.29	−	−	+	−	−
VP				−	−	−	−	−	64	1	5.48	−	−	+	−	−
VV				−	−	−	−	−	70	1	5.70	−	−	+	−	−
YD	37	1	5.45	−	+	−	−	+				−	−	−	−	−
YP				−	−	−	−	−	64	1	8.44	−	−	+	−	−

[1] L in peptide sequences represents Leu or Ile.

[2] Denono score was determined by PEAKS Studio during automatic *de novo* sequencing.

[3] z represents the charge state of the precursor ions of the identified peptides.

[4] α- and γ-forms of glutamyl peptides were determined by using a_1 and b_1 ions in the tandem MS spectra as diagnostic ions.

Table 7.3. Amino acid sequence and characteristics of the small peptides (2–5 amino acids) identified in common bean aquafaba through *de novo* sequencing.

Peptide[1]	Dimethylated							Nonderivatized						
	Denovo[2] score	z[3]	RT (min)	a_1 or b_1	Black bean	White bean	White kidney bean	Denovo score	z	RT (min)	a_1 or b_1	Black bean	White bean	White kidney bean
ADLF	51	1	28.17		−	+	−					−	−	−
AEL	39	1	13.21		+	+	+	70	1	11.38		−	+	−
AFL	37	1	26.02		−	+	−	33	1	24.35		−	−	+
ALK	70	2	4.92		−	+	+					−	−	−
ALL	40	1	23.57		−	+	+					−	−	−
AMP	41	1	13.03		−	+	−					−	−	−
AVL	50	1	17.42		+	+	+	64	1	15.78		−	−	+
AVP	44	1	11.72		+	+	+					−	−	−
DLL	52	1	24.82		−	+	−					−	−	−
DLQLK					−	−	−	62	2	17.05	a_1	−	+	−
γ-EACA[4]	84	1	4.81	b_1	+	−	−					−	−	−
α-EALE	63	1	11.44	a_1	−	+	−					−	−	−
γ-EC(Me)	50	1	5.50	b_1	+	+	+	80	1	4.30	b_1	+	+	+
γ-EC(Me)A	55	1	6.26	b_1	+	+	+	60	1	5.28	no a_1/b_1	+	+	−
γ-EC(Me)C(Me)	39	1	10.72	b_1	−	−	+	39	1	9.31	a1/b1	−	−	+
γ-EC(Me)EC(Me)	98	1	12.82	b_1	+	+	+	93	1	11.25	b_1	+	+	+
γ-EC(Me)EEC(Me)	96	1	13.98	b_1	−	−	+	96	1	12.36	b_1	−	+	+
γ-EC(Me)EL	99	1	21.79	b_1	+	+	+	98	1	19.98	tiny b_1	+	+	+

γ-EC(Me)EM	99	1	16.18	b_1	−	−	+					−	−	−
γ-EC(Me)F	77	1	21.72	b_1	+	+	+	76	1	19.91	tiny b_1	+	+	+
γ-EC(Me)L 1	80	1	17.60	b_1	+	+	−	78	1	16.11	b_1/a_1	+	−	−
γ-EC(Me)L 2	60	1	18.76	b_1	+	+	−					−	−	−
γ-EC(Me)R	78	1/2	4.56	b_1	+	+	+	69	2	2.62	b_1	+	−	+
γ-EC(Me)V	56	1	12.56	b_1	+	−	−					−	−	−
γ-EC(Me)W	77	1	24.07	b_1	+	+	+	77	1	22.75	b_1	+	+	+
EC(Me)Y					−	−	−	67	1	12.94	no a_1/b_1	−	−	+
γ-ECA	45	1	4.62	b_1	+	−	−					−	−	−
γ-ECC(Me)	41	1	11.53	b_1	+	−	−					−	−	−
γ-EDV	49	1	10.37	b_1	−	−	+					−	−	−
γ-EEC(Me)	75	1	7.37	b_1	+	+	+	64	1	6.34	b_1/a_1	+	+	+
γ-EEEC(Me)	99	1	8.63	b_1	+	+	+	98	1	7.54	b_1/a_1	+	+	+
γ-EEL	80	1	15.81	b_1	+	+	+	83	1	14.44	b_1/a_1	+	+	+
EEM					−	−	−	67	1	9.28	b_1/a_1	−	−	+
EEV					−	−	−	69	1	9.50	b_1/a_1	−	−	+
γ-EF	79	1	15.68	b_1	+	+	+	82	1	13.55	b1	+	+	+
α-EL	42	1	11.20	a_1	+	+	+	73	1	9.04	a1	−	+	+
γ-EL	55	1	13.84	b_1	+	+	+	85	1	11.88	b1	+	+	+
γ-ELA	78	1	13.10	b_1	+	−	−					−	−	−
γ-ELEC(Me)	99	1	19.67	b_1	+	+	+	99	1	17.77	tiny b_1	+	+	+
γ-ELEL	99	1	28.73	b_1	+	+	+	98	1	26.81	b_1	+	+	+
γ-ELEM	99	1	23.16	b_1	−	−	+	99	1	21.26	a_1/b_1	−	−	+

γ-ELF	76	1	30.69	b_1	+	+	−	72	1	29.29	no a_1/b_1	+	+	+
γ-ELP	78	1	20.24	b_1	−	−	+					−	−	−
γ-ELR	68	1/2	9.31	b_1	+	−	−	78	2	7.19	b_1	+	+	+
γ-EM	51	1	8.65	b_1	+	+	+	76	1	7.23	b_1/a_1	+	+	+
EMEC(Me)					−	−	−	91	1	13.79	no a_1/b_1	+	−	−
α-ENL	49	1	11.94	a_1	−	+	−					−	−	−
α-ENLA	60	1	11.14	a_1	−	+	−					−	−	−
ESW					−	−	−	71	1	15.97	no a_1/b_1	−	−	+
γ-EV	48	1	8.71	b_1	+	−	−	63	1	6.97	b_1	−	−	+
EVW					−	−	−	72	1	24.15	no a_1/b_1	−	−	+
γ-EW	75	1	18.96	b_1	+	+	+	76	1	16.99	tiny b_1	+	+	+
γ-EY					−	−	−	83	1	8.44	b_1	+	+	+
EYP					−	−	−	44	1	13.36	a1	−	−	+
FALN	71	1	23.81		+	−	−					−	−	−
FP	32	1	18.41		−	−	+					−	−	−
FP					−	−	−	66	1	15.64		−	−	+
FTTG	32	1	10.66		−	+	−					−	−	−
FVD	35	1	12.50		−	+	−	59	1	12.18		−	−	+
FVDA	56	1	14.35		+	−	−	92	1	14.29		+	+	−
FYFN	88	1	27.80		+	+	+					−	−	−
GDVF	87	1	20.87		−	+	−					−	−	−
GF	36	1	12.06		−	+	+					−	−	−
GFL 1	38	1	24.98		−	+	−	50	1	25.33		−	−	+

GFL 2	38	1	25.81	−	+	+					−	−	−
GFLQ	91	1	21.24	−	+	−					−	−	−
GFVP	91	1	23.68	−	−	+					−	−	−
GGF	27	1	12.04	−	+	−					−	−	−
GGFVP	78	1	24.28	−	+	+					−	−	−
GL	41	1	9.88	−	+	+	59	1	7.77		−	−	+
GLF	33	1	25.94	−	−	+					−	−	−
GLF	37	1	26.81	−	+	−					−	−	−
GLL 1	71	1	21.84	−	−	+					−	−	−
GLL 2	40	1	23.12	−	+	−					−	−	−
GLLGL	86	1	32.54	+	+	+					−	−	−
GLNNL	72	1	20.78	−	+	−					−	−	−
GLNP	82	1	13.11	−	+	+					−	−	−
GLP 1	46	1	15.37	−	+	+					−	−	−
GLP 2	47	1	16.46	−	+	+					−	−	−
GLV	37	1	16.49	−	+	−					−	−	−
GTLP	32	1	16.60	−	−	+					−	−	−
GTVVL	67	1	24.30	−	+	−					−	−	−
GVL	50	1	16.68	−	−	+					−	−	−
GVNLP	60	1	23.18	−	+	−					−	−	−
KL	76	2	6.00	−	−	+					−	−	−
KLL	77	2	15.14	+	+	+					−	−	−
LAL				−	−	−	66	1	22.09		−	+	+

LDL	34	1	24.95	−	−	+	67	1	22.25	−	−	+
LEC(Me)				−	−	−	70	1	12.26	+	+	+
LEL				−	−	−	77	1	21.73	+	−	+
LF 1	35	1	20.47	−	+	−	73	1	19.52	−	+	+
LF 2	35	1	22.44	−	+	−	67	1	21.31	−	+	−
LKLQ	85	2	16.08	−	+	−				−	−	−
LL 1	41	1	18.12	−	−	+	75	1	15.50	−	+	−
LL 2	42	1	20.04	−	+	+				−	−	−
LLEP	88	1	19.27	−	+	−	95	1	18.74	−	+	−
LLLQ				−	−	−	95	1	22.76	−	+	−
LLSGN				−	−	−	99	1	10.07	+	−	−
LNL				−	−	−	58	1	21.68	−	−	+
LP	56	1	10.77	−	+	+	62	1	9.55	−	+	−
LP				−	−	−	70	1	10.66	+	−	−
LQLK				−	−	−	66	2	13.45	−	+	−
LRLN				−	−	−	96	2	13.26	+	+	−
LSL				−	−	−	68	1	22.49	−	−	+
LVK	70	2	5.04	−	−	+				−	−	−
LVL	38	1	26.34	−	−	+	51	1	25.50	−	+	−
LVP	36	1	18.26	−	+	−				−	−	−
ME	38	1	16.34	+	−	−				−	−	−
NLM	38	1	16.58	−	−	+				−	−	−
NPLF	94	1	26.53	+	+	+	95	1	25.78	−	+	+

PALG				−	−	−	81	1	10.60	+	+	+
PFYFN				−	−	−	96	1	29.87	−	+	−
PGF				−	−	−	61	1	14.19	+	+	−
PGFP	83	1	20.59	+	+	+	90	1	20.72	+	+	−
PGFPN	77	1	16.49	−	+	+	96	1	16.75	+	−	−
PGLTL	81	1	26.93	−	−	+				−	−	−
PLE	47	1	9.38	−	+	−	46	1	8.67	−	+	−
PLF	48	1	25.70	+	+	+	49	1	25.20	+	+	+
PLFSD	83	1	22.29	+	+	+	91	1	22.18	−	−	+
PTL				−	−	−	59	1	14.09	−	−	+
PVL	38	1	17.53	−	+	−				−	−	−
SEGLN	52	1	9.08	−	+	−				−	−	−
SF	43	1	12.18	+	−	+				−	−	−
SFL				−	−	−	44	1	24.92	−	−	+
SFP	44	1	19.05	−	−	+				−	−	−
SL	41	1	10.01	−	+	+	71	1	6.28	−	−	+
SLF	38	1	25.63	−	+	+	48	1	25.78	−	−	+
SLF	35	1	26.98	−	+	−				−	−	−
SLL 1	52	1	21.82	−	+	+				−	−	−
SLL 2	41	1	23.42	−	+	+	72	1	21.73	−	+	−
SLLGR	90	2	13.66	−	+	−				−	−	−
SLM	44	1	18.67	−	+	−				−	−	−
SLMR	31	2	11.03	−	+	−				−	−	−

SLN	58	1	5.90	−	−	+				−	−	−
SLP	75	1	15.72	−	+	+				−	−	−
SLSL	91	1	24.65	+	+	+	93	1	22.69	−	+	−
SNLVP	52	1	21.84	−	+	+				−	−	−
SVF	34	1	20.26	−	+	−				−	−	−
SVL	38	1	16.88	−	−	+	70	1	15.31	−	+	−
SVTY	83	1	15.36	−	+	−				−	−	−
TF	42	1	15.78	−	+	+				−	−	−
TL	41	1	12.98	−	+	+				−	−	−
TLGN	70	1	6.92	+	+	+				−	−	−
TLL	50	1	24.22	−	+	−				−	−	−
TLNFK				−	−	−	76	2	14.89	+	−	−
TLTP	86	1	15.15	−	+	−				−	−	−
TMFA	66	1	22.71	−	+	−				−	−	−
TSF	33	1	15.76	−	+	−				−	−	−
TSLR				−	−	−	96	2	5.38	+	−	−
TVL	50	1	18.66	−	+	+				−	−	−
TVP	44	1	14.30	−	+	+				−	−	−
VAF				−	−	−	67	1	19.21	−	−	+
VAGL				−	−	−	68	1	17.07	−	−	+
VAK	69	2	4.45	−	+	−				−	−	−
VAV	45	1	14.21	−	+	−				−	−	−
VAVL	63	1	24.22	+	−	−	86	1	22.14	−	−	+

VAVLG	52	1	21.16	−	+	−				−	−	−
VDL				−	−	−	74	1	17.70	−	−	+
VF	36	1	16.49	−	−	+				−	−	−
VGL				−	−	−	36	1	15.56	−	−	+
VGL				−	−	−	62	1	16.55	−	−	+
VL	41	1	14.12	−	−	+				−	−	−
VLL 1	40	1	24.46	−	+	−	51	1	22.52	−	+	−
VLL 2	40	1	26.08	−	+	−	75	1	24.24	−	−	+
VNLP	49	1	23.72	−	−	+				−	−	−
VP				−	−	−	65	1	6.12	−	+	+
VQF				−	−	−	68	1	19.22	−	−	+
VSL				−	−	−	73	1	17.70	−	−	+
VTF	35	1	22.42	−	+	+	69	1	19.98	−	+	+
VTFL	53	1	32.34	−	+	+	96	1	30.68	−	+	+
VTL	39	1	20.70	−	+	+				−	−	−
VVVH	98	2	8.31	−	+	−	97	2	6.03	−	+	−
YDEP	40	1	11.67	−	−	+				−	−	−
YP				−	−	−	63	1	8.96	−	+	−

[1] L in peptide sequences represents Leu or Ile.

[2] Denono score was determined by PEAKS Studio during automatic *de novo* sequencing.

[3] z represents the charge state of the precursor ions of the identified peptides.

[4] α- and γ-forms of glutamyl peptides were determined by using a_1 and b_1 ions in the tandem MS spectra as diagnostic ions.

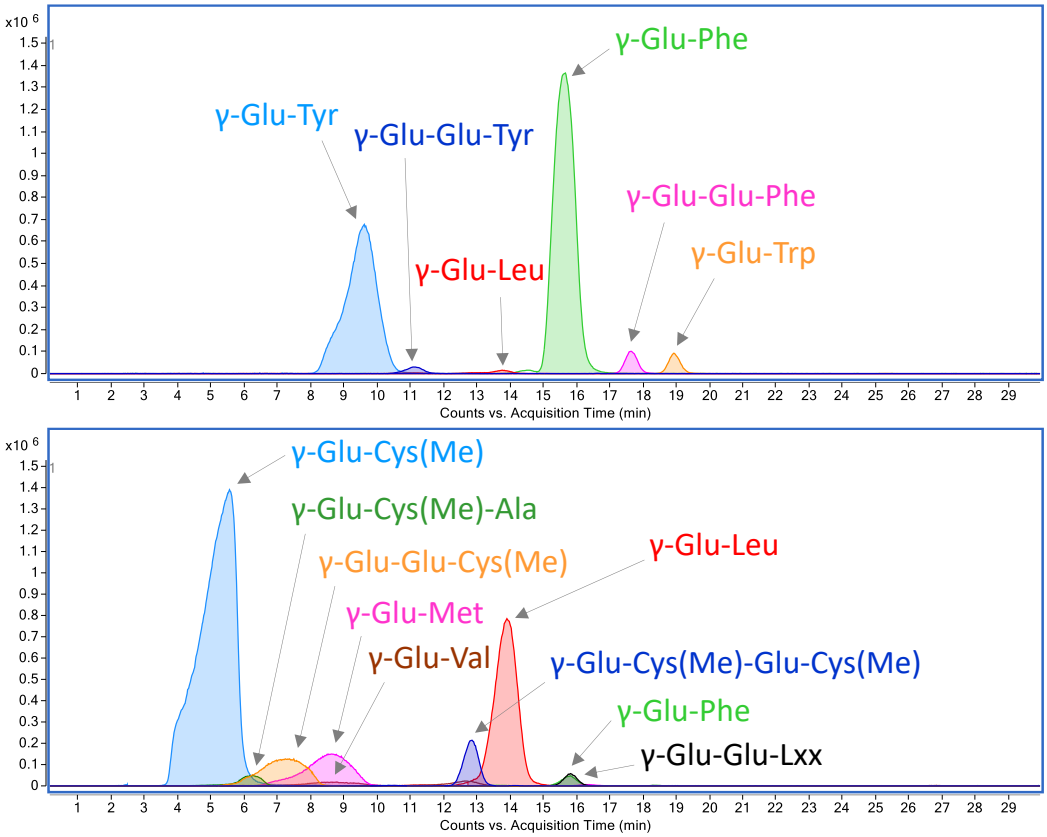


Fig. 7.8. LC-Q-TOF extracted ion chromatograms (EIC) of major γ-glutamyl peptides (dimethyl labeled) in aquafaba from chickpeas (upper panel; data of brand A) and common beans (bottom panel; data of white kidney beans). Cys(Me): S-methyl-Cys; Lxx: Leu or Ile.

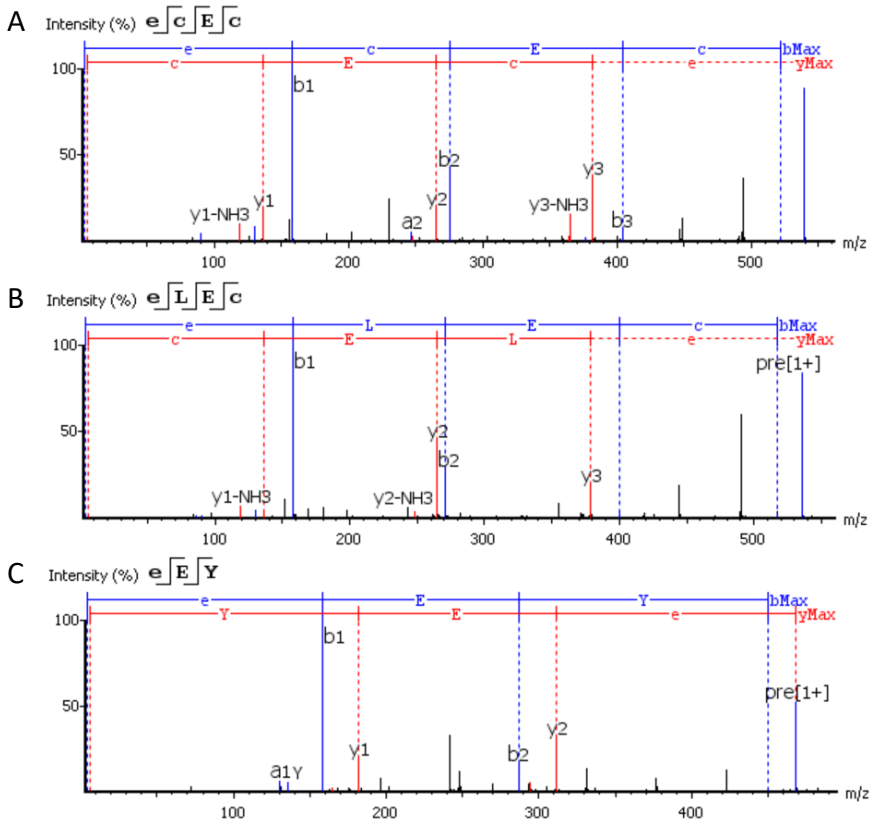


Fig. 7.9. Example tandem MS spectra of novel γ-glutamyl peptides identified in aquafaba from common beans (A, γ-Glu-S-methyl-Cys-Glu-S-methyl-Cys; B, γ-Glu-Lxx-Glu-S-methyl-Cys) and chickpeas (C, γ-Glu-Glu-Tyr) analyzed in the dimethyl labeled form. The significant b_1 ion peaks in the spectra confirmed the existence of an N-terminal γ-glutamine in the peptide structures. Lxx represents Leu or Ile.

7.3.2.2.3. Potential bioactivities of small peptides

Systematic searches against the bioactive peptide database BIOPEP-UWM (Minkiewicz et al., 2019) revealed that 34 and 32 small α-peptides identified (corresponding to 41 and 40 bioactive peptide sequences when taking all possible Leu/Ile-containing peptide sequences into account) in the aquafaba from chickpeas and common beans, respectively, possess specific bioactivities that were reported in previous studies (Table 7.4). The main identified activities included inhibition of

the angiotensin-converting enzyme I (ACE) and dipeptidyl peptidase IV (DPP IV), which are related to ameliorating cardiovascular disease. Because BIOPEP-UWM mainly contains information on α-peptides, a targeted literature search was further performed to explore the potential bioactivities of γ-glutamyl peptides. According to the literature, several γ-glutamyl peptides possess the unique ability to provide sensory and functional characteristics to the foods that contain them. For example, γ-Glu-Phe, γ-Glu-Tyr, γ-Glu-Met, γ-Glu-Val, and γ-Glu-Cys-β-Ala were reported to elicit a sensory phenomenon that is now known as the "kokumi sensation" (Amino et al., 2016; Dunkel et al., 2007; Shibata et al., 2017; Toelstede et al., 2009; Yamamoto et al., 2020; Yang et al., 2017; Yang, Sun-Waterhouse, Cui, Zhao, et al., 2018). When ingested alone, kokumi-inducing substances are often tasteless, yet they are able to induce the sensations of thickness, continuity, and mouthfeel when they are combined with other basic taste compounds, such as sodium chloride and umami substances (Dunkel et al., 2007; Li et al., 2020). Besides legumes, γ-glutamyl peptides are also found in garlic, onion, soy sauce (Yang et al., 2019), and ripened cheese, where they were demonstrated to play a critical role in providing the cheese's complex taste and long-lasting mouthfeel (Toelstede et al., 2009). Based on this limited but promising evidence, we see a clear path for the food industry to extract γ-glutamyl peptides from aquafaba and utilize them in the development of new foods, such as plant-based cheese-like products.

Additional bioactivities of γ-glutamyl peptides were also reported in the literature, such as anti-inflammatory activity (Guha et al., 2020; Zhang et al., 2015; Zhu et al., 2019) and DPP IV inhibition (Yang, Sun-Waterhouse, Cui, Dong, et al., 2018). In terms of mechanisms of action, it has been proposed that both the kokumi property and anti-inflammatory activity of γ-glutamyl peptides are associated with the allosteric activation of the calcium-sensing receptor (CaSR) (Guha

& Majumder, 2019, 2022; Guha et al., 2020; Ohsu et al., 2010), which is expressed on the cell surface of various tissues in the human body, including parathyroid glands, kidneys, liver, heart, gastrointestinal tract, and taste buds (Lu et al., 2021; Ohsu et al., 2010). Therefore, the current documentation from the existing literature indicates that γ-glutamyl peptides might be associated with other physiological functions by acting as CaSR agonists.

Table 7.4. Bioactivity identification for small α-peptides in chickpea and common bean aquafaba.[1]

Peptide[2]	Activities
Chickpea	
AEL	ACE inhibitor
AI	ACE inhibitor
AIP	ACE inhibitor
AL	Dipeptidyl peptidase IV inhibitor
ALP	ACE inhibitor
AVL	ACE inhibitor
EI	ACE inhibitor, dipeptidyl peptidase IV inhibitor
EL	Antioxidative
FL	Dipeptidyl peptidase IV inhibitor, dipeptidyl peptidase III inhibitor
FP	ACE inhibitor, dipeptidyl peptidase IV inhibitor
GF	ACE inhibitor, dipeptidyl peptidase IV inhibitor, dipeptidyl peptidase III inhibitor
GI	ACE inhibitor, dipeptidyl peptidase IV inhibitor
GL	ACE inhibitor, dipeptidyl peptidase IV inhibitor
GLP	ACE inhibitor
GY	ACE inhibitor, dipeptidyl peptidase IV inhibitor
IE	ACE inhibitor
IP	ACE inhibitor, dipeptidyl peptidase IV inhibitor
IV	Stimulating
LP	Dipeptidyl peptidase IV inhibitor
LV	Stimulating, dipeptidyl peptidase IV inhibitor
MG	ACE inhibitor, dipeptidyl peptidase IV inhibitor
NL	Dipeptidyl peptidase IV inhibitor
PGL	ACE inhibitor
SF	ACE inhibitor, dipeptidyl peptidase IV inhibitor, renin inhibitor
SI	Dipeptidyl peptidase IV inhibitor
SL	Dipeptidyl peptidase IV inhibitor
SV	Dipeptidyl peptidase IV inhibitor
TF	ACE inhibitor, dipeptidyl peptidase IV inhibitor, dipeptidyl peptidase III inhibitor, renin inhibitor
TI	Dipeptidyl peptidase IV inhibitor

TL	Dipeptidyl peptidase IV inhibitor
TT	Dipeptidyl peptidase IV inhibitor
TV	Dipeptidyl peptidase IV inhibitor
VD	Dipeptidyl peptidase IV inhibitor
VF	ACE inhibitor, dipeptidyl peptidase IV inhibitor
VGL	Dipeptidyl peptidase IV inhibitor
VI	Dipeptidyl peptidase IV inhibitor
VL	Stimulating, dipeptidyl peptidase IV inhibitor
VP	ACE inhibitor, dipeptidyl peptidase IV inhibitor
VV	Dipeptidyl peptidase IV inhibitor
YD	Dipeptidyl peptidase IV inhibitor
YP	ACE inhibitor, dipeptidyl peptidase IV inhibitor, alpha-glucosidase inhibitor
Common bean	
AEL	ACE inhibitor
AFL	ACE inhibitor
AVL	ACE inhibitor
AVP	ACE inhibitor
EI	ACE inhibitor, dipeptidyl peptidase IV inhibitor
EL	Antioxidative
FP	ACE inhibitor, dipeptidyl peptidase IV inhibitor
GF	ACE inhibitor, dipeptidyl peptidase IV inhibitor, dipeptidyl peptidase III inhibitor
GFL	Immunostimulating, regulating, dipeptidyl peptidase III inhibitor
GI	ACE inhibitor, dipeptidyl peptidase IV inhibitor
GL	ACE inhibitor, dipeptidyl peptidase IV inhibitor
GLF	Immunostimulating, regulating
GLP	ACE inhibitor
IF	ACE inhibitor
II	Stimulating, dipeptidyl peptidase IV inhibitor
IL	ACE inhibitor, dipeptidyl peptidase IV inhibitor, stimulating
IP	ACE inhibitor, dipeptidyl peptidase IV inhibitor
KI	Dipeptidyl peptidase IV inhibitor
KL	ACE inhibitor
LF	ACE inhibitor
LI	Stimulating, dipeptidyl peptidase IV inhibitor
LL	Stimulating, dipeptidyl peptidase IV inhibitor
LP	Dipeptidyl peptidase IV inhibitor
LVL	ACE inhibitor
ME	ACE inhibitor, dipeptidyl peptidase IV inhibitor
SF	ACE inhibitor, dipeptidyl peptidase IV inhibitor, renin inhibitor
SI	Dipeptidyl peptidase IV inhibitor
SL	Dipeptidyl peptidase IV inhibitor
SVL	Antioxidative
TF	ACE inhibitor, dipeptidyl peptidase IV inhibitor, dipeptidyl peptidase III inhibitor, renin inhibitor
TI	Dipeptidyl peptidase IV inhibitor
TL	Dipeptidyl peptidase IV inhibitor
VAF	ACE inhibitor

VAV	ACE inhibitor
VF	ACE inhibitor, dipeptidyl peptidase IV inhibitor
VGL	Dipeptidyl peptidase IV inhibitor
VI	Dipeptidyl peptidase IV inhibitor
VL	Stimulating, dipeptidyl peptidase IV inhibitor
VP	ACE inhibitor, dipeptidyl peptidase IV inhibitor
YP	ACE inhibitor, dipeptidyl peptidase IV inhibitor, alpha-glucosidase inhibitor

[1] Bioactivities were identified by searching against the BIOPEP-UWM bioactive peptide database (Minkiewicz et al., 2019).

[2] All possible bioactive peptide sequences for Leu/Ile-containing peptides were reported.

7.4. Conclusions

This study comprehensively characterized the oligosaccharides and peptides present in the aquafaba from chickpeas and common beans with optimized analytical workflows based on liquid chromatography and mass spectrometry. The workflow enabled the identification of oligosaccharides with various structures, including $Hex_{3–7}$, $Hex_{2–4}Pent_1$, $Hex_{2–3}HexOAc_1$, and $Hex_{2–4}Glycerol_1$. The main difference between the composition of aquafaba obtained from chickpeas and common beans was that ciceritol and other methyl-inositol-containing oligosaccharides were exclusively identified in chickpea aquafaba. For the analysis of small peptides, the incorporation of dimethyl labeling not only increased the number of small peptide identification but also enabled the differentiation of γ-glutamyl peptides from α-peptides. In particular, abundant a_1 and b_1 ions in the tandem MS spectra could be used as diagnostic ions for identifying α- and γ-glutamyl peptides. Overall, γ-glutamyl peptides accounted for a significant proportion of all the small peptides found in aquafaba. The compositions of γ-glutamyl peptides were substantially different between the chickpea and common bean aquafaba. Several γ-glutamyl peptides consisting of S-methyl-Cys were exclusively identified in common bean aquafaba. Besides the previously reported structures, new γ-glutamyl peptides were discovered in both

chickpea (e.g., γ-Glu-Glu-Tyr) and common bean (e.g., γ-Glu-S-methyl-Cys-Glu-S-methyl-Cys) aquafaba. Due to the unique ability of γ-glutamyl peptides to elicit kokumi sensation and activate CaSR, aquafaba could be potentially used as a natural taste-enhancing agent while reducing sodium content in food products. The discovery of bioactive peptides and oligosaccharides' presence in aquafaba can promote the development and functionality of novel plant-based products and improve the sustainability of food systems while diverting biomolecules from low-cost streams into value-added ingredients that improve human health.

Supporting informaion

Table 7.S1. Oligosaccharides identified in chickpea and common bean aquafaba.

Oligosaccharide	Retention time (min)	Observed m/z of oligosaccharide ions on ESI spectra				Chickpea					Common bean		
		$[M+H]^+$	$[M+NH_4]^+$	$[M+Na]^+$	$[M+K]^+$	Brand A	Brand B	Brand C	Brand D	Pressure cooker	Black	White	White kidney
Neutral oligosaccharides													
Hex2	0.707	343.122	**360.149**[1]	365.105	381.079	+	+	+	+	+	+	+	+
Hex1Methyl-inositol1	0.910	357.143	**374.165**	379.119	395.093	+	+	+	+	+	–	–	–
Hex2	1.184	343.123	**360.150**	365.105	381.078	+	+	+	+	+	+	+	+
Hex1Methyl-inositol1	1.414	357.138	**374.165**	379.121	395.094	+	+	+	+	+	–	–	–
Hex2	1.817	343.127	**360.149**	365.106	381.082	+	+	+	+	+	+	+	+
Hex2	2.305	343.126	**360.149**	365.106	381.079	+	+	+	+	+	+	+	+
Hex3	2.564	505.166	**522.202**	527.159	543.134	–	+	+	–	+	+	+	+
Hex2	3.353		**360.149**	365.103	381.080	–	–	–	–	–	+	–	–
Hex3	3.841	505.169	**522.200**	527.158	543.133	–	+	+	–	+	+	+	+
Hex2Methyl-inositol1	3.919	519.191	**536.218**	541.172	557.147	+	+	+	+	+	–	–	–
Hex2Glycerol1	4.514	417.159	**434.186**	439.142	455.115	+	+	+	+	+	+	+	+
Hex3	4.908	505.177	**522.203**	527.158	543.133	+	+	+	+	+	+	+	+
Hex3	5.916	505.176	**522.203**	527.157	543.134	+	+	+	+	+	+	+	+
Hex3	6.621	505.175	**522.204**	527.159	543.131	+	–	–	–	+	+	+	–

Hex3	7.529	505.176	**522.202**	527.158	543.134	+	+	+	+	+	+	+	+
Hex2Methyl-inositol1	7.749	519.194	**536.220**	541.174	557.147	+	+	+	+	+	–	–	–
Hex3	7.932	505.175	**522.203**	527.156	543.131	+	+	–	–	+	–	–	–
Hex2Pent1	8.981		**492.193**	497.145	513.117	+	–	–	–	–	+	+	+
Hex3	9.024	505.178	**522.202**	527.157	543.133	+	+	+	+	+	+	+	+
Hex3	9.528	505.177	**522.203**	527.158	543.131	+	+	+	+	+	–	–	–
Hex1HexOAc1	9.788		**402.160**	407.116	423.091	+	–	–	–	+	+	+	+
Hex3	9.886	505.177	**522.202**	527.160		–	–	–	–	–	+	+	+
Hex2Methyl-inositol1 (ciceritol)	10.103	519.194	**536.220**	541.174	557.147	+	+	+	+	+	–	–	–
Hex1HexOAc1	10.292		**402.160**	407.115	423.089	–	–	–	–	–	+	+	+
Hex4	10.832	667.228	**684.254**	689.211	705.185	+	+	+	+	+	+	+	+
Hex3Glycerol1	10.838	579.212	**596.239**	601.195	617.168	+	+	+	+	+	+	+	+
Hex3 (raffinose)	11.336	505.176	**522.204**	527.158	543.132	+	+	+	+	+	+	+	+
Hex4	11.700	667.226	**684.255**	689.212	705.188	+	+	+	+	+	+	+	+
Hex2Pent1	12.005	475.170	**492.194**	497.149	513.120	+	+	–	–	+	+	+	+
Hex5	12.106	829.278	**846.309**	851.261	867.234	+	+	+	+	+	+	+	+
Hex3	12.148	505.175	**522.203**	527.157	543.131	+	+	+	+	+	+	+	+
Hex3Methyl-inositol1	12.445	681.245	**698.273**	703.227	719.202	+	+	+	+	+	–	–	–
Hex4 (stachyose)	12.646	667.228	**684.258**	689.212	705.185	+	+	+	+	+	+	+	+
Hex3	13.055	505.176	**522.202**	527.127	543.129	+	+	–	+	–	–	–	+
Hex3HexOAc1	13.416		**726.263**	731.221	747.189	+	+	+	+	+	+	+	+
Hex3Methyl-inositol1	13.553	681.246	**698.269**	703.226	719.198	+	+	+	+	+	–	–	–

Hex4Glycerol1	13.660	741.269	**758.293**	763.248	779.219	+	+	+	+	+	+	+	+
Hex4	13.832	667.228	**684.254**	689.211	705.180	+	+	+	+	+	–	+	–
Hex5	14.366	829.281	**846.307**	851.263	867.231	+	+	+	+	+	+	+	+
Hex4Methyl-inositol1	14.668	843.300	**860.321**	865.275	881.247	+	+	+	+	+	–	–	–
Hex4Pent1	14.727		**816.298**	821.246		–	–	+	–	+	+	+	+
Hex5	14.769	829.275	**846.307**	851.263	867.233	+	+	+	+	+	+	+	+
Hex3Pent1	14.808	637.214	**654.243**	659.201	675.174	–	–	–	–	–	+	+	+
Hex3Methyl-inositol1	15.000	681.240	**698.266**	703.226	719.206	+	+	+	+	+	–	–	–
Hex4	15.071	667.228	**684.255**	689.210	705.184	+	+	+	+	+	+	+	+
Hex4	15.273	843.297	**860.321**	865.278		+	+	+	+	+	–	–	–
Hex5 (verbascose)	15.374	829.282	**846.309**	851.263	867.233	+	+	+	+	+	+	+	+
Hex4	15.474	667.228	**684.256**	689.210	705.181	+	+	+	+	+	–	+	+
Hex3HexOAc1	15.533		**726.268**			+	+	+	+	+	+	+	+
Hex5	15.676	829.283	**846.307**	851.263	867.235	+	+	+	+	+	+	+	+
Hex6	15.956	991.321	**1008.363**	1013.311	1029.277	+	+	+	+	+	+	–	+
Hex5	15.978	829.283	**846.308**	851.262	867.250	+	+	+	+	+	+	+	+
Hex4Methyl-inositol1	16.079	843.295	**860.323**	865.274	881.257	+	+	+	+	+	–	–	–
Hex6	16.359	991.334	**1008.358**	1013.305	1029.259	+	+	+	+	+	+	+	+
Hex5	16.463	829.280	**846.309**	851.263	867.237	+	+	+	+	+	+	+	+
Hex4	16.661	667.227	**684.257**	689.209	705.172	+	+	+	+	+	–	+	–
Hex7	16.785		**1170.410**	1175.354		+	+	+	+	+	+	+	+
Hex5	16.805	829.282	**846.308**	851.261	867.235	+	–	–	–	–	–	+	+

Hex3HexOAc1	16.944		**726.264**	731.218	747.188	+	+	+	+	+	+	+	+
Hex5	16.964	829.289	**846.307**	851.264	867.233	+	+	+	+	+	+	+	+
Hex6	16.964	911.145	**1008.359**	1013.311	1029.267	+	+	+	+	+	+	+	+
Hex4Pent1	17.246		**816.302**	821.247		−	−	+	−	+	+	+	+
Hex7	17.289	1153.399	**1170.414**	1175.362		+	+	+	+	+	+	+	+
Hex4	17.703	667.227	**684.255**	689.211	705.183	+	+	+	+	+	+	+	+
Hex3HexOAc1	17.773		**726.263**	731.221	747.173	+	+	+	+	+	+	+	+
Hex3HexOAc1	18.053		**726.264**	731.219	747.189	+	+	+	+	+	+	+	+
Hex6	18.498	991.333	**1008.357**	1013.317		+	+	+	+	+	+	+	+
Hex3	18.711	505.171	**522.201**	527.157	543.134	+	+	+	+	+	+	+	+
Hex5	18.720	829.289	**846.308**	851.269	867.218	+	+	+	+	+	+	+	+
Hex6	18.801	991.326	**1008.366**	1013.322	1029.296	+	+	+	+	+	+	+	+
Hex4	19.014		**684.256**	689.218		+	+	+	+	+	+	+	+
Hex3	19.182	505.178	**522.202**	527.154	543.128	+	+	+	+	+	+	+	+
Hex2HexOAc1	19.268		**564.216**	569.173	585.143	+	+	+	+	+	+	+	−
Hex4	26.107	667.228	**684.257**	689.209	705.186	+	+	+	+	+	−	−	−
Hex4	26.477	667.231	**684.255**	689.212	705.173	+	+	+	+	+	−	−	−
Charged oligosaccharides													
Hex3P1	5.342	585.143	**602.170**	607.123	623.094	−	+	−	−	+	+	+	+
Hex3P1	9.525	**585.141**	602.171	607.124	623.098	−	+	−	−	+	+	+	+
Hex3P1	10.103	**585.141**		607.124	623.098	−	+	−	−	+	+	+	+
Hex3P1	10.432	585.143	**602.166**	607.123	623.093	−	+	−	−	+	+	+	+
Hex4P1	12.599	**747.195**		769.177	785.148	+	+	+	−	+	+	+	+

Hex4P1	13.204	**747.198**	764.223	769.177	785.152	+	+	+	+	+	+	+	+
Hex4P1	14.647	**747.195**	764.220	769.177	785.150	+	+	+	−	+	+	+	+
Hex4P1	15.957	**747.194**		769.176	785.145	+	+	+	−	+	+	+	+
Hex4P1	12.751	747.195	**764.222**	769.176	785.145	+	+	+	−	+	+	+	+
Hex4P1	14.061	747.191	**764.215**	769.177	785.144	+	+	+	−	+	+	+	+

[1] Oligosaccharide ions with m/z values marked in bold were the precursors for tandem MS identification.

References

Afshin, A., Micha, R., Khatibzadeh, S., & Mozaffarian, D. (2014). Consumption of nuts and legumes and risk of incident ischemic heart disease, stroke, and diabetes: a systematic review and meta-analysis. *The American Journal of Clinical Nutrition*, *100*(1), 278–288. https://doi.org/10.3945/ajcn.113.076901

Amino, Y., Nakazawa, M., Kaneko, M., Miyaki, T., Miyamura, N., Maruyama, Y., & Eto, Y. (2016). Structure-CaSR-Activity Relation of Kokumi γ-Glutamyl Peptides. *Chemical & Pharmaceutical Bulletin*, *64*(8), 1181–1189. https://doi.org/10.1248/cpb.c16-00293

Becerra-Tomás, N., Papandreou, C., & Salas-Salvadó, J. (2019). Legume consumption and cardiometabolic health. *Advances in Nutrition (Bethesda, Md.)*, *10*(Suppl_4), S437–S450. https://doi.org/10.1093/advances/nmz003

Borges, I. F., Barbedo, C. J., Richter, A. A., & Figueiredo-Ribeiro, R. de C. L. (2006). Variations in sugars and cyclitols during development and maturation of seeds of brazilwood (Caesalpinia echinata Lam., Leguminosae). *Brazilian Journal of Plant Physiology*, *18*(4), 475–482. https://doi.org/10.1590/S1677-04202006000400005

Díaz-Batalla, L., Widholm, J. M., Fahey, G. C., Castaño-Tostado, E., & Paredes-López, O. (2006). Chemical components with health implications in wild and cultivated Mexican common bean seeds (Phaseolus vulgaris L.). *Journal of Agricultural and Food Chemistry*, *54*(6), 2045–2052. https://doi.org/10.1021/jf051706l

Dunkel, A., Köster, J., & Hofmann, T. (2007). Molecular and sensory characterization of gamma-glutamyl peptides as key contributors to the kokumi taste of edible beans (Phaseolus vulgaris L.). *Journal of Agricultural and Food Chemistry*, *55*(16), 6712–6719. https://doi.org/10.1021/jf071276u

Elango, D., Rajendran, K., Van der Laan, L., Sebastiar, S., Raigne, J., Thaiparambil, N. A., El Haddad, N., Raja, B., Wang, W., Ferela, A., Chiteri, K. O., Thudi, M., Varshney, R. K., Chopra, S., Singh, 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

Food and Agriculture Organization of the United Nations (FAO). (2022). *ESS: Crops Statistics - Concepts, Definitions and Classifications*. https://www.fao.org/economic/the-statistics-division-

ess/methodology/methodology-systems/crops-statistics-concepts-definitions-and-classifications/en/

Gangola, M. P., Jaiswal, S., Khedikar, Y. P., & Chibbar, R. N. (2014). A reliable and rapid method for soluble sugars and RFO analysis in chickpea using HPAEC-PAD and its comparison with HPLC-RI. *Food Chemistry*, *154*, 127–133. https://doi.org/10.1016/j.foodchem.2013.12.085

Guha, S., & Majumder, K. (2019). Structural-features of food-derived bioactive peptides with anti-inflammatory activity: A brief review. *Journal of Food Biochemistry*, *43*(1), e12531. https://doi.org/10.1111/jfbc.12531

Guha, S., & Majumder, K. (2022). Comprehensive Review of γ-Glutamyl Peptides (γ-GPs) and Their Effect on Inflammation Concerning Cardiovascular Health. *Journal of Agricultural and Food Chemistry*. https://doi.org/10.1021/acs.jafc.2c01712

Guha, S., Paul, C., Alvarez, S., Mine, Y., & Majumder, K. (2020). Dietary γ-Glutamyl Valine Ameliorates TNF-α-Induced Vascular Inflammation via Endothelial Calcium-Sensing Receptors. *Journal of Agricultural and Food Chemistry*, *68*(34), 9139–9149. https://doi.org/10.1021/acs.jafc.0c04526

He, Y., Meda, V., Reaney, M. J. T., & Mustafa, R. (2021). Aquafaba, a new plant-based rheological additive for food applications. *Trends in Food Science & Technology*, *111*, 27–42. https://doi.org/10.1016/j.tifs.2021.02.035

Hsu, J.-L., & Chen, S.-H. (2016). Stable isotope dimethyl labelling for quantitative proteomics and beyond. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, *374*(2079). https://doi.org/10.1098/rsta.2015.0364

Huang, Y.-P., Dias, F. F. G., Leite Nobrega de Moura Bell, J. M., & Barile, D. (2022). A complete workflow for discovering small bioactive peptides in foods by LC-MS/MS: A case study on almonds. *Food Chemistry*, *369*, 130834. https://doi.org/10.1016/j.foodchem.2021.130834

Huang, Y.-P., Robinson, R. C., & 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., & 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

Ismail, B. P., Senaratne-Lenagala, L., Stube, A., & Brackenridge, A. (2020). Protein demand: review of plant and animal proteins used in alternative protein product development and production. *Animal Frontiers (Online)*, *10*(4), 53–63. https://doi.org/10.1093/af/vfaa040

Liao, D., Cram, D., Sharpe, A. G., & Marsolais, F. (2013). Transcriptome Profiling Identifies Candidate Genes Associated with the Accumulation of Distinct Sulfur γ-Glutamyl Dipeptides in Phaseolus vulgaris and Vigna mungo Seeds. *Frontiers in Plant Science*, *4*, 60. https://doi.org/10.3389/fpls.2013.00060

Lioi, L. (1989). Variation of the storage protein phaseolin in common bean (Phaseolus vulgaris L.) from the Mediterranean area. *Euphytica*, *44*(1–2), 151–155. https://doi.org/10.1007/BF00022610

Li, Q., Zhang, L., & Lametsch, R. (2020). Current progress in kokumi-active peptides, evaluation and preparation methods: a review. *Critical Reviews in Food Science and Nutrition*, 1–12. https://doi.org/10.1080/10408398.2020.1837726

Lu, Y., Wang, J., Soladoye, O. P., Aluko, R. E., Fu, Y., & Zhang, Y. (2021). Preparation, receptors, bioactivity and bioavailability of γ-glutamyl peptides: A comprehensive review. *Trends in Food Science & Technology*, *113*, 301–314. https://doi.org/10.1016/j.tifs.2021.04.051

Machida, K., Huang, Y.-P., Furlan Gonçalves Dias, F., Barile, D., & Leite Nobrega de Moura Bell, J. M. (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

Mekky, R. H., Contreras, M. del M., El-Gindi, M. R., Abdel-Monem, A. R., Abdel-Sattar, E., & Segura-Carretero, A. (2015). Profiling of phenolic and other compounds from Egyptian cultivars of chickpea (Cicer arietinum L.) and antioxidant activity: a comparative study. *RSC Adv.*, *5*(23), 17751–17767. https://doi.org/10.1039/C4RA13155J

Minkiewicz, P., Iwaniak, A., & Darewicz, M. (2019). BIOPEP-UWM Database of Bioactive Peptides: Current Opportunities. *International Journal of Molecular Sciences*, *20*(23). https://doi.org/10.3390/ijms20235978

Morris, Clayton J., Thompson, J. F., & Zacharius, R. M. (1963). The Identification of γ-l-Glutamyl-l-leucine and γ-l-Glutamyl-l-methionine in Kidney Bean Seeds (Phaseolus vulgaris). *Journal of Biological Chemistry*, *238*(2), 650–652. https://doi.org/10.1016/S0021-9258(18)81313-8

Morris, C J, & Thompson, J. F. (1962). The isolation and characterization of gamma-L-glutamyl-L-tyrosine and gamma-L-glutamyl-L-phenylalanine from soybeans. *Biochemistry*, *1*, 706–709. https://doi.org/10.1021/bi00910a026

Mustafa, R., He, Y., Shim, Y. Y., & Reaney, M. J. T. (2018). Aquafaba, wastewater from chickpea canning, functions as an egg replacer in sponge cake. *International Journal of Food Science & Technology*, *53*(10), 2247–2255. https://doi.org/10.1111/ijfs.13813

Mustafa, R., & Reaney, M. J. T. (2020). Aquafaba, from Food Waste to a Value-Added Product. In R. Campos-Vega, B. D. Oomah, & H. A. Vergara-Castañeda (Eds.), *Food wastes and by-products: nutraceutical and health potential* (pp. 93–126). Wiley. https://doi.org/10.1002/9781119534167.ch4

Ohsu, T., Amino, Y., Nagasaki, H., Yamanaka, T., Takeshita, S., Hatanaka, T., Maruyama, Y., Miyamura, N., & Eto, Y. (2010). Involvement of the calcium-sensing receptor in human taste perception. *The Journal of Biological Chemistry*, *285*(2), 1016–1022. https://doi.org/10.1074/jbc.M109.029165

Oku, S., Ueno, K., Tsuruta, Y., Jitsuyama, Y., Suzuki, T., Onodera, S., Maeda, T., & Shimura, H. (2019). Sugar accumulation and activities of enzymes involved in fructan dynamics from seedling to bulb formation in onion (Allium cepa L.). *Scientia Horticulturae*, *247*, 147–155. https://doi.org/10.1016/j.scienta.2018.12.013

Paizs, B., & Suhai, S. (2005). Fragmentation pathways of protonated peptides. *Mass Spectrometry Reviews*, *24*(4), 508–548. https://doi.org/10.1002/mas.20024

Quemener, B., & Brillouet, J.-M. (1983). Ciceritol, a pinitol digalactoside form seeds of chickpea, lentil and white lupin. *Phytochemistry*, *22*(8), 1745–1751. https://doi.org/10.1016/S0031-9422(00)80263-0

Ruiz-Aceituno, L., Carrero-Carralero, C., Ruiz-Matute, A. I., Ramos, L., Sanz, M. L., & Martínez-Castro, I. (2017). Characterization of cyclitol glycosides by gas chromatography coupled to mass spectrometry. *Journal of Chromatography. A*, *1484*, 58–64. https://doi.org/10.1016/j.chroma.2017.01.001

Saboori-Robat, E., Joshi, J., Pajak, A., Solouki, M., Mohsenpour, M., Renaud, J., & Marsolais, F. (2019). Common Bean (Phaseolus vulgaris L.) Accumulates Most S-Methylcysteine as Its γ-Glutamyl Dipeptide. *Plants*, *8*(5). https://doi.org/10.3390/plants8050126

Schoeninger, V., Coelho, S. R. M., & Bassinello, P. Z. (2017). Industrial processing of canned beans. *Ciência Rural*, *47*(5). https://doi.org/10.1590/0103-8478cr20160672

Shibata, M., Hirotsuka, M., Mizutani, Y., Takahashi, H., Kawada, T., Matsumiya, K., Hayashi, Y., & Matsumura, Y. (2017). Isolation and characterization of key contributors to the "kokumi" taste in soybean seeds. *Bioscience, Biotechnology, and Biochemistry*, *81*(11), 2168–2177. https://doi.org/10.1080/09168451.2017.1372179

Siva, N., Thavarajah, P., & Thavarajah, D. (2020). Prebiotic carbohydrate concentrations of common bean and chickpea change during cooking, cooling, and reheating. *Journal of Food Science*, *85*(4), 980–988. https://doi.org/10.1111/1750-3841.15066

Stagnari, F., Maggio, A., Galieni, A., & Pisante, M. (2017). Multiple benefits of legumes for agriculture sustainability: an overview. *Chemical and Biological Technologies in Agriculture*, *4*(1), 2. https://doi.org/10.1186/s40538-016-0085-1

Tavano, O. L., & Neves, V. A. (2008). Isolation, solubility and in vitro hydrolysis of chickpea vicilin-like protein. *LWT - Food Science and Technology*, *41*(7), 1244–1251. https://doi.org/10.1016/j.lwt.2007.08.003

Taylor, M., Chapman, R., Beyaert, R., Hernández-Sebastià, C., & Marsolais, F. (2008). Seed storage protein deficiency improves sulfur amino acid content in common bean (Phaseolus vulgaris L.): redirection of sulfur from gamma-glutamyl-S-methyl-cysteine. *Journal of Agricultural and Food Chemistry*, *56*(14), 5647–5654. https://doi.org/10.1021/jf800787y

Toelstede, S., Dunkel, A., & Hofmann, T. (2009). A series of kokumi peptides impart the long-lasting mouthfulness of matured Gouda cheese. *Journal of Agricultural and Food Chemistry*, *57*(4), 1440–1448. https://doi.org/10.1021/jf803376d

USA Pulses. (2022). *Processing Information and Technical Manual - Chapter 5: Applications*. https://www.usapulses.org/technical-manual/chapter-5-applications/canning

Wang, N., Hatcher, D. W., Tyler, R. T., Toews, R., & Gawalko, E. J. (2010). Effect of cooking on the composition of beans (Phaseolus vulgaris L.) and chickpeas (Cicer arietinum L.). *Food Research International*, *43*(2), 589–594. https://doi.org/10.1016/j.foodres.2009.07.012

Wang, W., Wright, E. M., Uebersax, M. A., & Cichy, K. (2021). A pilot-scale dry bean canning and evaluation protocol. *Journal of Food Processing and Preservation*. https://doi.org/10.1111/jfpp.16171

Xiaoli, X., Liyi, Y., Shuang, H., Wei, L., Yi, S., Hao, M., Jusong, Z., & Xiaoxiong, Z. (2008). Determination of oligosaccharide contents in 19 cultivars of chickpea (Cicer arietinum L) seeds by high performance liquid chromatography. *Food Chemistry*, *111*(1), 215–219. https://doi.org/10.1016/j.foodchem.2008.03.039

Yamamoto, M., Terada, Y., Motoyama, T., Shibata, M., Saito, T., & Ito, K. (2020). N-terminal [Glu]3 moiety of γ-glutamyl peptides contributes largely to the activation of human calcium-sensing receptor, a kokumi receptor. *Bioscience, Biotechnology, and Biochemistry*, *84*(7), 1497–1500. https://doi.org/10.1080/09168451.2020.1743169

Yang, J., Bai, W., Zeng, X., & Cui, C. (2019). Gamma glutamyl peptides: The food source, enzymatic synthesis, kokumi-active and the potential functional properties – A review. *Trends in Food Science & Technology*, *91*, 339–346. https://doi.org/10.1016/j.tifs.2019.07.022

Yang, J., Sun-Waterhouse, D., Cui, C., Dong, K., & Wang, W. (2017). Synthesis and Sensory Characteristics of Kokumi γ-[Glu]n-Phe in the Presence of Glutamine and Phenylalanine: Glutaminase from Bacillus amyloliquefaciens or Aspergillus oryzae as the Catalyst. *Journal of Agricultural and Food Chemistry*, *65*(39), 8696–8703. https://doi.org/10.1021/acs.jafc.7b03419

Yang, J., Sun-Waterhouse, D., Cui, C., Dong, K., & Zhao, M. (2018). *γ* -Glu-Met synthesised using a bacterial glutaminase as a potential inhibitor of dipeptidyl peptidase IV. *International Journal of Food Science & Technology*, *53*(5), 1166–1175. https://doi.org/10.1111/ijfs.13692

Yang, J., Sun-Waterhouse, D., Cui, C., Zhao, H., & Dong, K. (2018). Gamma-glutamylation of the white particulates of sufu and simultaneous synthesis of multiple acceptor amino acids-containing γ -glutamyl peptides: Favorable catalytic actions of glutaminase. *LWT*, *96*, 315–321. https://doi.org/10.1016/j.lwt.2018.05.055

Zhang, H., Kovacs-Nolan, J., Kodera, T., Eto, Y., & Mine, Y. (2015). γ-Glutamyl cysteine and γ-glutamyl valine inhibit TNF-α signaling in intestinal epithelial cells and reduce inflammation in a mouse model of colitis via allosteric activation of the calcium-sensing receptor. *Biochimica et Biophysica Acta*, *1852*(5), 792–804. https://doi.org/10.1016/j.bbadis.2014.12.023

Zhu, X., Tao, Q., Sun-Waterhouse, D., Li, W., Liu, S., & Cui, C. (2019). γ-[Glu]n-Trp ameliorates anxiety/depression-like behaviors and its anti-inflammatory effect in an animal model of anxiety/depression. *Food & Function*, *10*(9), 5544–5554. https://doi.org/10.1039/c9fo01467e

Chapter VIII

Effects of proteolysis on almond protein profile, digestibility, and antigenicity

(This chapter is a manuscript in preparation; authors: Dias, F.F.G., **Huang, Y.-P.**, Schauer J., Barile, D., Van de Water, J., & de Moura Bell, J.M.L. Huang, Y.-P. conducted the proteomic analysis and wrote the original draft of the corresponding sections of the manuscript.)

Abstract

Almonds (*Prunus dulcis*) are one of the most consumed tree nuts worldwide and have been recognized as a healthy and nutritious food. Nevertheless, almonds are also a source of allergenic proteins that can trigger several mild to life-threatening immunoreactions. The effects of selected enzymatic extraction conditions on the protein profile, determined by proteomics of excised SDS-PAGE gel bands, *in vitro* protein digestibility, and immunoreactivity of almond protein extracts were evaluated. Proteolysis altered almond protein sequential and conformational characteristics thus affecting their digestibility and antigenicity. Proteomics revealed that enzymatic extraction resulted in the complete hydrolysis of Prunin 1 and 2 α-subunits with higher resistence to hydrolysis of Prunin 1 and 2 β-subunits and a reduction of allergen proteins and epitopes. Protein *in vitro* digestibility increased from 79.1 to 88.5% after proteolysis, as determined by a static digestion model. The degree of hydrolysis and peptide content of the enzymatically extracted proteins during gastric and duodenal digestion was significantly higher than the ones from the unhydrolyzed protein. Proteolysis resulted in a 75% reduction in almond protein immunoreactivity as determined by a sandwich enzyme-linked immunosorbent assay and reduction in IgE and IgG reactivities using human sera. The present study shows that moderated hydrolysis using protease can be used as a strategy to improve almond protein digestibility and reduce antigenicity. Our findings could further enhance the potential use of almond protein hydrolysates in the formulation of hypoallergenic products with enhanced nutritional quality and safety.

Keywords: protein hydrolysis, proteomics, almond protein *in vitro* digestibility, allergenicity

8.1. Introduction

The growing demand for plant-based protein sources has been driven by the need to feed an increasing world population with sustainable and nutritious foods. To that end, plant-based protein ingredients that rival or have improved functional and biological properties (e.g., improved digestibility and reduced allergenicity) compared to the ones from traditional animal protein ingredients must be developed (Akharume et al., 2021).

Tree nuts (e.g., almonds, walnuts, and cashews, among others) are a good source of high-quality protein and lipids, ranking high among the healthiest snacks (Geiselhart et al., 2018). Despite their dense nutritional content, three nuts are one of the eight food groups accounting for the majority of food-induced allergies, with their consumption being associated with several mild to life-threatening immunoreactions in sensitive groups (Sicherer et al., 2003; Tiwari et al., 2010).

Almonds *(Prunus dulcis*) are one of the most produced tree nuts in the world with a forecast production of 1.6 million metric tons for 2021/2022 (on a shelled basis) (USDA, 2021). They are also one of the most consumed tree nuts worldwide, being highly appreciated for their pleasant taste and abundance in nutritional compounds (lipids, high-quality proteins, vitamin E, and polyphenols) (Sathe, 1993; Yada, Lapsley, & Huang, 2011) and ease of application in a wide range of products (i.e., snacks, dairy alternatives, gluten-free flours), being particularly attractive as a source of protein for vegetarian and vegan diets (Tomishima et al., 2021). However, the desirable techno-functional, nutritional, and texture properties of almonds that allow such applications are highly dependent on the almond protein characteristics (Dias & de Moura Bell, 2022; Wolf & Sathe, 1998). Proteins are of great importance in food processing and product development as they impart many of the functional and nutritional properties that can drive consumers' acceptance of the product. Despite the attractive properties of almond proteins, almond-induced allergies are the

third most commonly reported tree nut allergy in the United States, with a prevalence of 0.7% in the population (Gupta et al., 2019). Therefore, the development of processing strategies to improve almond protein quality and utilization is of great interest.

Environmentally friendly strategies such as aqueous and enzymatic aqueous extraction processes have been used to simultaneously extract lipids, proteins, and soluble carbohydrates from almond flour, avoiding the upstream use of mechanical pressing and/or flammable solvent extraction to produce defatted flours for protein extraction (Dias et al., 2020, 2022; Dias & de Moura Bell, 2022). While the benefits of using enzymes to assist the extraction have been evaluated regarding the overall extractability of oil and protein from almond flour and almond cake (Almeida et al., 2019; Dias et al., 2020; Souza et al., 2019) and the functional properties of the almond protein (Dias and Bell, 2022; Amirshaghaghi, Rezaei, & Rezaei, 2017; Sze-Tao & Sathe, 2000), the impact of enzymatic extraction on the digestibility and allergenicity of almond protein has yet to be evaluated

Because proteolysis can entail protein structural modifications that might alter their functional, nutritional, and biological properties (de Souza et al., 2020a, Dias et al., 2020, 2022; Dias & de Moura Bell, 2022), in agreement with the common use of proteolysis to produce hypoallergenic dietary products from different protein sources, the overall goal of this study was to determine the effects of proteolysis, arising from the enzymatic extraction (EAEP) of full-fat almond flour, on the protein molecular weight profile of excised gel bands by liquid chromatography-mass spectrometry (LC-MS/MS), *in vitro* protein digestibility (total protein digestibility, degree of hydrolysis and peptide quantification kinetics), and almond protein allergen quantification (Sandwich ELISA) and antigenicity (IgE and IgG Western blotting). The elucidation of the impact of sustainable flammable solvent-free extraction methods (i.e., aqueous

vs. enzyme-assisted aqueous extraction) on the digestibility and allergenicity of the extracted proteins is critical to the development of a bio-guided process that will deliver high-quality and safe food ingredients for subsequent applications.

8.2. Materials and methods

8.2.1. Materials

Almond flour (obtained from a mix of *Prunus dulcis* varieties) was kindly provided by Blue Diamond Growers (Sacramento, CA, USA). Whole almonds were blanched and sieved through a US#12 mesh (1.70 mm sieve size) (ultra-fine granulometry), with a minimum recovery of 85%. The sample D [4,3] was 245 µm and the D (10), D (50), and D (90) were 0.4, 146, and 714 µm, respectively (Mastersizer 3000E - Malvern Panalytical Inc., Westborough, MA, USA). The almond flour proximate composition was 42.6 ± 0.6 % of oil, 27.9 ± 0.8% of carbohydrates, 21.7 ± 0.6% of protein, 5.3 ± 0.1% of moisture, 2.4 ± 0.1% of ash. Moisture, fat, and ash were determined according to AOAC methods 925.09, 989.05, and 920.125, respectively (AOAC, 1990). Protein content was determined by the Dumas combustion method using a conversion factor of 5.18 (Vario MAX cube, Elementar Analysensysteme GmbH, Germany). Carbohydrates were determined by difference (100 − the sum of other components) (Ghribi et al., 2015). Each analysis was performed in triplicate and data was reported as the mean ± standard deviation.

A neutral endoprotease from *Bacillus subtilis* (, 5.5 to 9.5 optimum pH range and 30 to 70 ^{o}C optimum temperature range, and 2×10^{6} PC/g of activity) was kindly supplied by Bio-Cat (Bio-Cat Inc., Virginia, NY, USA). Sodium dodecyl sulfate (SDS), casein, 1-anilino-8-naphthalenesulfonate (ANS) were acquired from VWR Inc. (Chicago, Illinois, USA). Pepsin from porcine gastric mucosa (3706 U/mg), pancreatin from porcine pancreas (100 U/mg), amylase from

porcine pancreas (1005 U/mg), mucin, bile salts, L-serine, and o-phthaldialdehyde (OPA), were purchased from Millipore Sigma (St. Louis, MO, USA). Bovine serum albumin was purchased from Thermo Fisher Scientific (Waltham, Massachusetts, USA). Soybean protein isolated powder was acquired in a local grocery store (Davis, CA, USA). Tris buffer, β-mercaptoethanol, Laemmli sample buffer, Coomassie Blue G250, and Dual-color standard (10-250 kDa) were purchased from Bio-Rad (Hercules, CA, USA). All other chemicals were of analytical grade.

8.2.2. Almond protein extraction methods

Almond protein extracts were produced by aqueous (AEP - unhydrolyzed) and enzymatic extraction processes (EAEP - hydrolyzed) from full-fat almond flour, as described by (Dias & de Moura Bell, 2022). For the AEP, 700 g of almond flour was dispersed into water to achieve a solids-to-liquid ratio of 1:10 (w/v) in a 10-L jacketed glass reactor (CG-1965-610M - Chemglass Life Sciences LLC, Vineland, NJ, USA). The extraction was performed at pH 9.0, 50 °C, for 60 min under constant stirring (120 rpm). For the EAEP, 0.5% (w/v) (weight of enzyme per weight of almond flour) of Neutral Protease was added to the slurry, and extractions were performed as described for the AEP. After the extraction, the slurry was centrifuged at 3,000 × g for 30 min at 25 °C to remove the insoluble fraction. The liquid fraction was placed back into the glass reactor and allowed to separate overnight at 4 °C into the protein-rich phase (protein extract) and oil-rich phase (cream). AEP and EAEP protein extracts were stored at -20 °C until subsequent analysis. Each extraction process was performed in triplicate. The proximate composition of AEP and EAEP protein extracts, determined as described in item 2.1, was 57.3 and 59.2% protein (d.b.), 8.1 and 7.1% lipids (d.b.), 12.2 and 9.7% ash (d.b.), 18.7 and 20.9% carbohydrates (d.b.), and 2.7 and 3.0% moisture,respectively.

8.2.3. Proteomics analysis of excised gel bands

8.2.3.1. Protein electrophoresis-based separation (SDS-PAGE)

Proteins from AEP and EAEP extracts were separated using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) in the presence of a reducing agent (β-mercaptoethanol) as described by Laemmli, 1970. Samples were extracted using Laemmli buffer (1:1, v/v) for 5 min at 95 °C. Samples were then cooled at room temperature and loaded onto a precast 12% acrylamide gel. Electrophoretic separation was carried out at 200 V at room temperature for 1 h. Coomassie Blue G250 was used to stain the gel and a dual-color standard (10-250 kDa) was used as the molecular weight marker. The gel image and polypeptide distribution for the protein gels were obtained using the Gel DOCTM EZ Imager system and Image Lab software (Bio-Rad, Hercules, CA, USA). Gel bands were cut into 9 sections for AEP and 5 sections for EAEP extracts as shown in Fig. 8.1A. The excised gel bands were then placed in a 1.7 mL tube containing 150 µL of RO water and stored at 4 °C until analysis.

8.2.3.2. Trypsin digestion

Each gel band section was diced into small pieces and placed in a 1.5 mL tube. In-gel digestion on the gel pieces was conducted as described by Gundry et al., 2009. Briefly, the gel pieces were destained with a water-methanol mixture (1:1, v/v), washed with water, and dehydrated with acetonitrile. The disulfide bonds were reduced with 100 µl of 10 mM dithiothreitol at 55 °C for 45 min; the free cysteines were alkylated by 100 µl of 55 mM iodoacetamide at room temperature for 30 min. The gel pieces were washed with 25 mM ammonium bicarbonate in 50% acetonitrile (v/v) and dehydrated with acetonitrile. Trypsin digestion was performed by adding 10 µg/mL trypsin prepared in 25 mM ammonium bicarbonate to cover the gel pieces, incubating at 4 °C for 1 h and then at 37 °C overnight. The released peptides were collected by extracting the gel

pieces with 50% acetonitrile, and 1% trifluoroacetic acid in water (v/v/v). The peptide sample was dried using a centrifugal evaporator (MiVac Quattro, Genevac Ltd., Ipswitch, Suffolk, UK).

8.2.3.3. Peptide sample cleanup

The tryptic peptide sample was re-dissolved in 0.1% trifluoroacetic acid in water (v/v) and loaded to a C18 solid-phase extraction column (Discovery DSC-18, 500 mg, 3 mL tube, MilliporeSigma, St. Louis, MO, USA) preconditioned with 5 mL acetonitrile followed by 5 mL 0.1% trifluoroacetic acid. The column was washed with 6 mL 0.1% trifluoroacetic acid. Peptides were recovered by flushing the column with 6 mL of an aqueous solution composed of 80% acetonitrile, and 0.1% trifluoroacetic acid in water (v/v/v). The purified peptide sample was dried using a centrifugal evaporator.

8.2.3.4. Liquid chromatography with tandem mass spectrometry (LC-MS/MS) analysis

The purified tryptic peptides were analyzed by LC-MS/MS on a Q Exactive Plus Orbitrap Mass spectrometer in conjunction with Proxeon Easy-nLC 1200 HPLC (Thermo Scientific, Waltham, MA, USA) and Proxeon nanospray source. A volume containing 1 µg of peptides was loaded onto a 100 µm × 25 mm Dr. Maisch 100Å 5U reverse-phase trap where the peptides were desalted online before being separated using a 75 µm × 150 mm Dr. Maisch 200Å 3U reverse-phase column. Peptides were eluted using an 80-min gradient with a flow rate of 300 nL/min. The mobile phase was composed of 0.1% formic acid in water (A) and acetonitrile (B). The gradient was programmed as follows: 0–48 min: 2–20% B; 48–60 min: 20–35% B; 60-62 min: 35–100% B; 62–64 min: 100% B; 64–65 min: 100–2% B; 65–80 min: 2% B. An MS survey scan was obtained for the m/z range 300–1600; MS/MS spectra were acquired using a top 15 method, where the top 15 ions in the MS spectra were subjected to HCD (High Energy Collisional Dissociation). An isolation mass window of 1.6 m/z was used for the precursor ion selection, and normalized

collision energy of 27% was used for fragmentation. A five-second duration was used for the dynamic exclusion.

8.2.3.5. Data analysis

The LC-MS/MS data were analyzed by PEAKS Studio X+ (Bioinformatics Solutions Inc., Waterloo, ON, Canada). Peptides and proteins were identified through database search using almond (*Prunus dulcis*) protein sequences downloaded from the UniProt database (https://www.uniprot.org/, accessed 3/10/2020), including both Swiss-Prot and TrEMBL. The error tolerance was 10.0 ppm and 0.02 Da for the precursor and fragment ions, respectively. Semispecific digestion using trypsin as the enzyme with three maximal missed cleavage was used for predicting the precursor peptides. The variable modifications included deamidation on asparagine and glutamine, phosphorylation on serine, threonine, and tyrosine, oxidation on methionine, and carbamidomethylation on cysteine; additional unspecific modifications and mutations were found by using the PEAKS PTM followed by the SPIDER function. Peptide identifications were filtered with the criterion of $-10\lg P \geq 35$ and protein identifications with $-10\lg P \geq 50$ as well as ≥ 5 unique peptides. Due to the existence of protein isoforms and homologous regions among different proteins in the UniProt protein database, manual curation was conducted to avoid redundant protein identifications by combining the proteins being identified mainly based on the same set of peptides into one protein identification.

8.2.4. In vitro protein digestibility

AEP and EAEP protein extracts were subjected to *in vitro* digestion to assess the impact of the extraction methods employed (aqueous vs. enzymatic extraction) on total protein digestibility and the effects of the simulated gastrointestinal digestion process on protein molecular weight profile, degree of hydrolysis, and peptide content. The *in vitro* simulated gastrointestinal digestion

was performed as described by Bornhorst & Singh, 2013 and de Souza, Dias, Oliveira, de Moura Bell, & Koblitz, 2020 using simulated saliva (SSF), gastric (SGF), and intestinal (SIF) fluids to mimic the oral-gastro-duodenal digestion. The composition of the simulated fluids is presented in Supplementary material Table 8.S1. Casein and soybean isolated protein powders were used for comparison purposes. For the oral phase, 5 mL of each sample was mixed with 3.33 mL of SSF and vortexed for 30 s. Subsequently, the simulated oral bolus was mixed with 6.66 mL of SGF. The pH was adjusted to 3.0 and the gastric digesta was incubated for 120 min at 37 °C and 120 rpm. The simulated gastric digesta was mixed with 10 mL of SIF. The solution pH was adjusted to 7.0 and the simulated digesta was incubated for 120 min at 37 °C and 120 rpm. To stop the digestion, samples were heated in a water bath at 85 °C for 3 min. A 24% (w/v) trichloroacetic acid (TCA) solution was added to the samples in a 1:1 (v/v) proportion and the samples were centrifuged at 4,000 rpm for 30 min at 4 °C. Total *in vitro* protein digestibility was calculated using Equation 1. Total nitrogen (NT) and nonprotein nitrogen (NPN - soluble fraction after TCA precipitation) were assessed by the Dumas combustion method (Vario MAX cube, Elementar Analysensysteme GmbH, Germany).

$$Protein\ digestibility\ (\%) = 100\ x\ \left(\frac{NPN_{after} - (NPN_{before} - NPN_{blank})}{NT_{before} - NPN_{before}}\right) \qquad \text{(Eq. 1)}$$

where: NPN_{after} = protein after digestion, NPN_{before}= protein before digestion, NPN_{blank} = enzyme blank and NT_{before} = total protein before digestion.

8.2.4.1. Protein molecular weight profile

SDS-PAGE was used to evaluate changes in the protein molecular weight of the almond samples due to oral, gastro, and duodenal digestion. Aliquots of AEP and EAEP samples were collected at 0, 0.5, 30, 60, 90, 120, 150, 180, 210, and 240 min of digestion and placed in a water bath at 85 °C for 3 min to stop the digestion process. The protein molecular weight profile was

assessed as described in item 2.3.1. The gel was imaged using a Gel Doc™ EZ Imager system and Image Lab software (Bio-Rad Laboratories, CA, USA).

8.2.4.2. Degree of hydrolysis

The degree of hydrolysis (DH) of the aliquots from the digestibility kinetics was evaluated by the o-phthaldialdehyde method (OPA) as described by Nielsen, Petersen, and Dambmann, 2001. Briefly, 400 uL of a 2% (w/v) solution was added to 3 mL of OPA reagent, the mixture was vortexed and let stand for 2 min at room temperature, and the absorbance was measured at 340 nm. An L-serine solution (0.9516 meqv/L) was used as standard. A blank solution was prepared with distilled water instead of sample and used as the reaction control. The protein percentage in the protein extracts was obtained by the Dumas method (conversion factor 5.18), with the equipment Vario MAX cube (Elementar Analysensysteme GmbH, Germany).

The DH was determined as follows:

$$\text{DH } (\%) = \frac{h}{h_{tot}}\ x\ 100 \qquad \text{(Eq. 2)}$$

Where h is the number of hydrolyzed bonds. h_{tot} is the total number of peptide bonds per protein equivalent (7.58) (Liu et al., 2016).

8.2.4.3. Peptide quantification

Aliquots from the digestion kinetics were precipitated using ice-cold ethanol (2:1) (ethanol:sample), incubated for 2 h at -20 °C, and centrifuged at 4000 x g for 30 min at 4 C. The supernatant was separated and used for the analysis. Briefly, 20 µL of diluted samples were pipetted in a 96-well plate followed by the addition of 200 uL of Fluoraldehyde™ o-phthaldialdehyde (OPA) reagent solution (Thermo Fisher Scientific-Waltham, MA, USA). A blank was made by adding water instead of sample to the solution. The microplate was agitated for 5 min in a shaker at 300 rpm. The sample's fluorescence was determined at 340 nm (excitation)

and 455 nm (emission) using a microplate reader (SpectraMax iD5 Multi-Mode Microplate Reader, Molecular Devices, San Jose, California, USA). A Bovine Serum Albumin standard curve (BSA) was prepared at 0, 4, 8, 16, 20, and 40 µg (r^2=0.995) and used for the peptide quantification. Samples were analyzed in triplicate.

8.2.5. Sandwich ELISA for almond immunoreactivity

Almond reactivity of AEP (unhydrolyzed) and EAEP (hydrolyzed) samples was initially determined by the Veratox kit for almond allergen (Neogen, Lansing, MI, USA). The samples were prepared as recommended by the manufacturer and a rabbit antibody-based inhibition sandwich ELISA assay was used for detecting and quantifying the presence of amandin (AMP), a major allergenic protein in almonds. An AMP calibration curve at concentrations of 0, 2.5, 5, 10, 15, 20 and 25 mg/L (r^2=0.9997) was used. The samples (n=6 replicates) were diluted to fall within the AMP standard curve and read at 450 nm in a microplate reader (SpectraMax iD5 Multi-Mode Microplate Reader, Molecular Devices, San Jose, California, USA).

8.2.6. Immunoreactivity by Western blotting

8.2.6.1. Initial screening

Five human blood (4C, 35C, 78C, 38, and 196b) sera from patients showing strong IgE reactivity to almonds were used for the initial screening. Sample BB12, from a patient showing no reactivity to almonds, was used as a negative control. AEP and EAEP samples (270 µg) were loaded in a 12% Bis/Tris preparative gel, which was run for 45 minutes before being transferred to a 0.45 µm nitrocellulose membrane according to the method described by Towbin et al., 1979. The five human sera samples previously known to be immunogenic to almonds were tested at three dilutions (1:10, 1:20, and 1:40) along with a control sample with no known allergy (BB12). The sera were incubated overnight with the nitrocellulose strips at room temperature. The strips were

then washed and incubated with mouse anti-Human IgE Fc HRP secondary antibody at 1:10,000 for detection. Those preliminary blots (Supplementary material Fig. 8.S1A, B) showed that only Human serum 4C sample 196b exhibited reactivity with bands above 60 kDa and below 20 kDa. Moreover, it was determined that a 1:20 dilution of human sera and a 1:5000 dilution (per Abcam's recommendation) of secondary antibody would be sufficient to show differences in the reactive potential of the protein samples.

8.2.6.2. IgE immunoblotting

Two samples were chosen to be run by Western blotting: 196b and 4C. In each of these blots, 26 µg of each almond extract (AEP and EAEP) was reduced and run on a 12% Bis/Tris 15 well gel for 45 minutes before being transferred to a nitrocellulose membrane. Sample 4C and 196b were used at a 1:20 dilution for IgE. The mouse anti-Human IgE Fc secondary antibody dilution was adjusted to 1:5000. AEP and EAEP samples were tested in triplicate (A1-3 for AEP and B1-3 for EAEP). Band relative quantification was performed using Image J (Schneider et al., 2012).

8.2.6.3. IgG immunoblotting

Samples 196b and 4C were again chosen to be tested by Western blot. In each of these blots, 26 µg of each almond extract was reduced and run on a 12% Bis/Tris 15 well gel for 45 minutes before being transferred to a nitrocellulose membrane. Samples 4C and 196b were used at a 1:200 dilution. Goat anti-Human IgG secondary antibody was used at a dilution of 1 to 10,000. AEP and EAEP samples were tested in triplicate (A1-3 for AEP and B1-3 for EAEP).

8.2.7. Statistical analysis

The results are given as the means ± one standard deviation. Data were analyzed in the Statistica™ Software (TIBCO Software Inc, Palo Alto, CA, US) using a one-way ANOVA and Tukey's post hoc with $p < 0.05$.

8.3. Results and discussion

8.3.1. Effects of enzymatic hydrolysis on the protein profile by proteomic analysis of the excised gel bands

8.3.1.1. Protein profiles

The SDS-PAGE protein and peptide gels (Fig. 8.1A) showed that the use of enzyme during the extraction significantly affected the composition and molecular weight profile of almond proteins. To better understand the impact of proteolysis on almond protein composition, LC-MS/MS-based proteomics analysis was carried out in order to identify the specific proteins of interest. Selected protein bands from the SDS-PAGE (Fig. 8.1A) were in-gel digested by trypsin enzyme, and the resulting peptide pool was analyzed.

Detailed information about the proteins identified (ranked by the total peak area of the tryptic peptides generated from each protein) from each gel cut, including protein accession, protein name, sequence coverage, total peak area, and relative abundance, was reported in Supplementary material Tables 8.S2–8.S15. Peak areas of the peptides belonging to the same protein identification were summed up for estimating the relative abundance of the identified proteins in each gel cut. Protein identifications including at least five unique peptides, and having a total peak area of above 1.0×10^6 or a sequence coverage above 25%, were reported in Tables 8.S2–8.S15.

Fig. 8.1B summarized the number of protein identifications and the major proteins (above 1% relative abundance) identified in the gel cut samples. For the AEP and EAEP samples, each gel cut included 19–95 and 6–24 protein identifications, respectively. The fewer proteins identified in each EAEP gel cut indicated that a significant portion of proteins was hydrolyzed into low-molecular-weight peptides by the enzyme used for assisting the extraction and that those small peptides were not captured in the SDS-PAGE gel.

Among the AEP gel cut samples, AEP 1 (~50–78 kDa), AEP 6 (~22–29 kDa), and AEP 9 (~9–18 kDa) had higher numbers of protein identifications (95, 63, and 86, respectively), likely because these gel samples contained different proteins with broader molecular weight range. Besides, the three gel cut samples only included less intensive bands, which might also enable the identification of low abundance proteins. Regarding the relative abundance of different proteins, prunin 1 was the most abundant protein in all gel cut samples, except for AEP 2, AEP 8, and EAEP 3, which had vicilin (predicted), prunin 2, and prunin 2, respectively, as the most abundant protein (Fig. 8.1B). Prunin 1 and prunin 2 are composed of two polypeptides (acidic α-subunit and basic β-subunit) that can then be assembled into a hexameric protein named amandin. In almonds, amandin accounts for ~65% of water-extractable proteins and is also considered the major almond protein allergen (Wolf & Sathe, 1998). Therefore, identifying ways to reduce amandin immunoreactivity is of great interest.

Prunin 1 and prunin 2 comprised a significant portion in most gel cuts (Fig. 8.1B), in agreement with the high abundance of amandin. To assist in ascertaining the protein constitutions in each gel cut, the total peak areas of the identified tryptic peptides originated from the α- and β-subunit regions of prunin 1 and prunin 2 (Garcia-Mas, Messeguer, Arús, & Puigdomènech, 1995) were calculated separately (Fig. 8.1C).

AEP 1 gel cut (~50–78 kDa) contained 66.0% of prunin 1 and 10.0% of prunin 2, which should be mainly the precursors of prunin 1 (61.0 kDa) and prunin 2 (55.9 kDa) with both α- and β-subunit regions linked. This is based on the fact that the α- and β-subunit regions were both identified in high abundance, whereas the signal peptide regions (prunin 1 (1–20) and prunin 2 (1–11)) (Garcia-Mas, Messeguer, Arús, & Puigdomènech, 1995) were not detected. AEP 3 (~39 kDa) and AEP 4 (~34 kDa) consisted of mostly prunin 1 (93.5 and 89.2%, respectively) and prunin 2 (2.9 and 7.1%, respectively), with a stronger contribution of the α-subunit regions. According to the above results and the molecular weight reported by Garcia-Mas, Messeguer, Arús, & Puigdomènech, 1995, prunin 1 α-subunit (40.1 kDa) seems to be the dominant protein constituent in AEP 3 and AEP 4, while prunin 2 α-subunit (34.5 kDa) was present yet in a small portion in AEP 4. For AEP 5 (~30 kDa), the high abundance of prunin 1 (86.4%), especially its α-subunit region (78.0%), indicated that this gel band might be associated with prunin 1 α-subunit, although the molecular weight was slightly lower than the literature value (40.1 kDa). AEP 7 (~21.5 kDa) contained 93.8% of prunin 1, mainly belonging to the β-subunit region. In comparison, AEP 8 (~19 kDa) included 78.6% of prunin 2 and 20.3% of prunin 1, pertaining to the β-subunit region. Therefore, AEP 7 and AEP 8 were mainly composed of prunin 1 β-subunit (20.9 kDa) and prunin 2 β-subunit (21.4 kDa), respectively. For AEP 2 (~46 kDa), AEP 6 (~22–29 kDa), and AEP 9 (~9–18 kDa), the sum of the percentages of prunin 1 and prunin 2 was near 15, 80, and 50%, respectively. The molecular weights of the three gel cut samples were dissimilar to either the precursor or the α- and β-subunits of prunin 1 and prunin 2. A previous study also showed that several minor bands, other than the above-mentioned major polypeptides, were observed on the SDS-PAGE gel of almond extracts (Sathe et al., 2002). The composition of the three samples' α- and β-subunit regions were close to AEP 1 which contained the precursor polypeptides (Fig. 8.1C),

suggesting that AEP 2, AEP 6, and AEP 9 may include the protein fragments of both prunin 1 and prunin 2 within the respective molecular weight ranges.

The EAEP gel cut samples contained 74.9–99.5% of prunin 1 and prunin 2, which is similar to the percentages in most of the AEP gel samples (76.0–96.4%, excluding AEP 2). However, all the EAEP gel samples were dominated by the β-subunit regions of prunin 1 and prunin 2 (Fig. 8.1C), demonstrating the higher resistance of the β-subunit against the proteolysis by the protease used. EAEP 1 (~25–35 kDa) consisted of some faint smeared bands, which included 63.8% of prunin 1 (2.6% α-subunit and 66.0% β-subunit) and 11.1% of prunin 2 (0.2% α-subunit and 10.7% β-subunit). Because the AEP's gel bands above 25 kDa were mostly not seen in the EAEP extracts, EAEP 1 should mainly contain partially hydrolyzed products from these proteins and any unhydrolyzed proteins that maintained their original molecular weight. Besides the high abundance of the β-subunit regions, EAEP 1 also contained some vicilins (predicted) with various sequences (Fig. 8.1B). Because the molecular weight of intact vicilin protein chains is much higher than the upper range of EAEP 1 (e.g., A0A5E4EZP4: 90.4 kDa; A0A5E4FV72: 57.6 kDa), the vicilins present in EAEP 1 should be hydrolyzed products of vicilin proteins. EAEP 2 (~21–22 kDa) contained 98.0% of the β-subunit region and possessed a similar molecular weight to prunin 1 β-subunit (20.9 kDa), indicating that EAEP 2 was mainly the intact prunin 1 β-subunit. EAEP 3 (~18.5–20 kDa) also consisted of mostly the β-subunit regions of prunin 1 (31.7%) and prunin 2 (59.6%), which were suggested to be the identities of the two major bands on EAEP 3. EAEP 4 (~15–17 kDa) and EAEP 5 (~9–14 kDa) had 77.9% and 97.1%, respectively, of the combination of Prunin 1 and prunin 2, with the majority belonging to the β-subunit region. Similar to the AEP 9, the molecular weight ranges of EAEP 4 and EAEP 5 samples were below the sizes of the precursor polypeptides and the α- and β-subunits of prunin 1 and prunin 2. Thus, these gel bands

were mainly the proteolyzed products from prunin 1 and prunin 2. Overall, in all the EAEP gel cut samples, only EAEP 4 and EAEP 5, representing the lower molecular weight range, contained a small portion of the α-subunit region of prunin 1. Prunin 2 α-subunit region only accounted for 0.0–0.6% of the abundance in the EAEP gel samples. These results reveal that the α-subunit regions were easier to be hydrolyzed by the neutral protease than the β-subunit regions. Despite prunin 1 α-subunit being observed in high abundance in the AEP samples, during EAEP, it was likely broken down into low-molecular-weight peptides that cannot be detected by the SDS-PAGE.

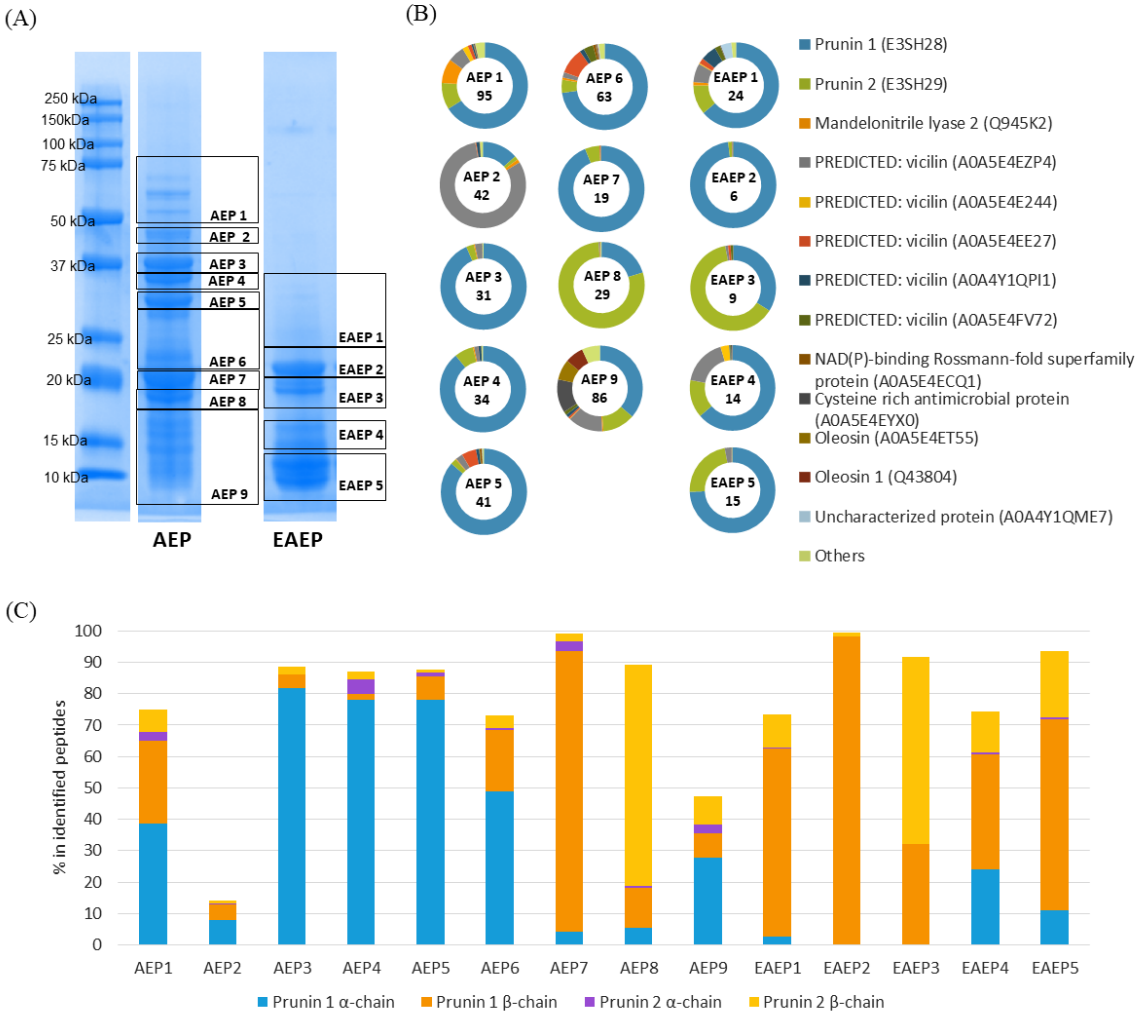


Fig. 8.1. SDS-PAGE protein profile of AEP and EAEP almond samples indicating the gel cuts for proteomics analysis (A). Relative abundance of protein identifications (numbers in the middle of pie charts

represent the number of protein identification) (B) and of α- and β-subunits in prunin 1 and prunin 2 (C) analyzed by LC-MS/MS-based proteomics analysis.

8.3.1.2. Identification of protein allergens

To date, several protein allergens in almonds, including Pru du 1, Pru du 2, Pru du 3, Pru du 4, Pru du 5, Pru du 6, Pru du 8, Pru du 10, and Pru du AP, have been reported (Supplementary material Table 8.S16). By comparing the UniProt accession numbers of the protein allergens with the proteomics data from the gel cut samples, protein allergens present in each gel cut were identified and listed in Table 8.1. Pru du 1, Pru du 4, Pru du 6, Pru du 8, Pru du 10, and Pru du AP were found in at least one gel cut. Among these, Pru du 1 and Pru du 4 were exclusively identified in AEP 9. As mentioned above, prunin 1 and prunin 2, which belong to the allergen Pru du 6, were found in all the gel cut samples with a significant sequence coverage. Pru du 10 (mandelonitrile lyase 2) and Pru du 8 (cysteine-rich antimicrobial protein) were the other two protein allergens found, which had ≥ 1% relative abundance in at least one gel cut sample. Because all the bands in AEP 1 through AEP 5 (~30–78 kDa) were almost imperceptible on the SDS-PAGE gel of EAEP, the proteins in this range, including the intact polypeptides and the α- and β-subunits of Pru du 6 as well Pru du 10, were likely broken down into smaller fragments by the neutral protease during EAEP, which could indicate a reduction in the antigenicity potential of this sample.

Mandelonitrile lyase 2 (Pru du 10), with a theoretical molecular weight of 60.0 kDa (for the mature protein without modifications) and four N-glycosylation sites (Dreveny et al., 2001), is a recently identified protein allergen (Kabasser et al., 2021) in almond. Dreveny et al. (2001) showed that mandelonitrile lyase 2 isolated from almond flour appeared as a single band on SDS-PAGE gel at ~60 kDa, which is similar to the theoretical value. Thus, mandelonitrile lyase 2 was expected to be found in the AEP 1 gel cut sample (~50–78 kDa). According to the proteomics

analysis, mandelonitrile lyase 2 accounted for 9.2% of the protein abundance in AEP 1, with a sequence coverage of 50%. Because the molecular weight of mandelonitrile lyase 2 is close to the molecular weight of prunin 1, it is difficult to annotate the exact band for mandelonitrile lyase 2 in the AEP 1 sample. However, the absence of bands on the EAEP gel in the same molecular weight region (~60 kDa) indicated the destruction of mandelonitrile lyase 2 during enzymatic extraction (EAEP). Mandelonitrile lyase 2 was also found in other AEP gel cut samples with a sequence coverage ranging from 23 to 43%, but the relative abundance in those samples was low (0.0–1.3%). The protein allergen was detected in some of the EAEP gel cut samples with a sequence coverage of 6–31%. The highest sequence coverage (31%) and the highest relative abundance (1.3%) were found in EAEP 1. As the bands in EAEP 1 were extremely faint, it can be concluded that mandelonitrile lyase 2 was drastically decreased by the neutral protease used in EAEP.

Cysteine-rich antimicrobial protein, which was initially speculated to be 2S albumin based on sequence similarity (Poltronieri et al., 2002) and later to be vicilin (Garino et al., 2015), was recently identified as a new family of allergen proteins (Che et al., 2019), named Pru du 8 in the WHO/IUIS database. Che et al. (2019) showed that the recombinant cysteine-rich antimicrobial protein band located at ~31 kDa on an SDS-PAGE gel. This protein was found to be present in the AEP 9 gel cut (~9–18 kDa) with a significant abundance (12.6%) and had the highest sequence coverage (74%; 84% when excluding the signal peptide sequence f(1–30)) (Supplementary material Fig. 8.S2) and number of identified peptides (54 sequences) than all the other gel cut samples. Although it was also identified in other AEP gel cuts (e.g., AEP 5 and AEP 6), the relative abundance was low (<0.5%). The primary location of cysteine-rich antimicrobial protein on an SDS-PAGE gel found in the present study appeared to agree with the band location of its natural

form at 12 and 13 kDa reported by Poltronieri et al., 2002 and Kabasser et al., 2021, respectively. Poltronieri et al., 2002 found that the isolated 12 kDa IgE-binding protein contained the N-terminal region of cysteine-rich antimicrobial protein, whereas Kabasser et al. (2021) identified the purified 13 kDa protein as a C-terminal fragment of the same protein. Interestingly, in the current study, peptide identifications for AEP 9 gel cut covered both the N- and C-terminal regions (Supplementary material Fig. 8.S2). The lower molecular weight of the natural protein compared with the recombinant protein might be related to uncharacterized post-translational proteolytic processing (Che et al., 2019). It is possible that the two previous studies (Poltronieri et al., 2002 and Kabasser et al., 2021) isolated the N- and C-terminal fragments, respectively, of cysteine-rich antimicrobial protein due to the use of different purification techniques. Because no prior purification was conducted in the current study, both fragments were able to be identified in the gel cut AEP 9. Cysteine-rich antimicrobial protein possibly contains several disulfide bonds, which might make the protein more resistant to proteolysis. Indeed, the cysteine-rich antimicrobial protein was detected in EAEP 4 and EAEP 5, which covered most of the AEP 9 molecular weight range. Nonetheless, the lower relative abundance (0.5% and 0.1%, respectively), sequence coverage (43% and 56%, respectively), and the number of peptides identified (21 and 19 sequences, respectively) in the EAEP 4 and EAEP 5 gel cuts, compared with AEP 9, suggested that, during the EAEP, the enzyme partially damaged the linear structure of this protein allergen. Overall, the use of protease to assist the extraction of almond proteins led to the reduction of the presence of allergen-related proteins, which could indicate a potential reduction in its antigenicity.

Table 8.1. Identification of allergen proteins in bands from SDS-PAGE of AEP and EAEP samples by LC-MS/MS analysis of tryptic peptides produced by in-gel protein digestion, coupled to commercial software-based protein identification.

Band	Accession	Protein	Allergens	MW (kDa)	AAs [1]	−10lgP[2]	Coverage (%)	Number of peptides
AEP 1	E3SH28	Prunin 1	Pru du 6.0101	63.1	551	496.20	83	232
	Q43607	Prunin 1	Pru du 6	63.0	551	493.85	80	225
	Q43608	Prunin 2 (Fragment)	Pru du 6	57.1	504	368.24	59	64
	E3SH29	Prunin 2 (Fragment)	Pru du 6, Pru du 6.0201	57.0	504	366.23	62	66
	Q945K2	Mandelonitrile lyase 2	Pru du 10	61.2	563	339.11	50	61
	P82952	γ-Conglutin 1	Pru du γ-conglutin	46.9	431	263.06	44	18
AEP 2	E3SH28	Prunin 1	Pru du 6.0101	63.1	551	462.80	70	138
	Q43607	Prunin 1	Pru du 6	63.0	551	460.18	73	141
	Q945K2	Mandelonitrile lyase 2	Pru du 10	61.2	563	306.37	36	31
	Q43608	Prunin 2 (Fragment)	Pru du 6	57.1	504	302.14	43	24
	E3SH29	Prunin 2 (Fragment)	Pru du 6, Pru du 6.0201	57.0	504	300.72	44	26
	P82952	γ-Conglutin 1	Pru du γ-conglutin	46.9	431	279.64	52	22
AEP 3	E3SH28	Prunin 1	Pru du 6.0101	63.1	551	585.21	86	360
	Q43607	Prunin 1	Pru du 6	63.0	551	573.74	86	356
	E3SH29	Prunin 2 (Fragment)	Pru du 6, Pru du 6.0201	57.0	504	358.60	59	54
	Q43608	Prunin 2 (Fragment)	Pru du 6	57.1	504	352.32	53	49
	P82952	γ-Conglutin 1	Pru du γ-conglutin	46.9	431	285.39	53	22
	Q945K2	Mandelonitrile lyase 2	Pru du 10	61.2	563	273.63	36	26
AEP 4	Q43607	Prunin 1	Pru du 6	63017	551	534.47	83	268
	E3SH28	Prunin 1	Pru du 6.0101	63052	551	533.61	83	267
	E3SH29	Prunin 2 (Fragment)	Pru du 6, Pru du 6.0201	57006	504	391.59	66	78
	Q43608	Prunin 2 (Fragment)	Pru du 6	57052	504	379.54	60	73
	Q945K2	Mandelonitrile lyase 2	Pru du 10	61158	563	259.22	35	22
	P82952	γ-Conglutin 1	Pru du γ-conglutin	46945	431	153.61	11	4
	A0A516F3L2	Cysteine-rich antimicrobial protein	Pru du 8, Pru du 8.0101	31068	264	171.86	31	7
AEP 5	E3SH28	Prunin 1	Pru du 6.0101	63052	551	577.89	89	350
	Q43607	Prunin 1	Pru du 6	63017	551	570.30	89	347
	E3SH29	Prunin 2 (Fragment)	Pru du 6, Pru du 6.0201	57006	504	367.77	68	63
	Q43608	Prunin 2 (Fragment)	Pru du 6	57052	504	364.50	62	60
	Q945K2	Mandelonitrile lyase 2	Pru du 10	61158	563	259.06	34	22

	A0A516F3L2	Cysteine-rich antimicrobial protein	Pru du 8, Pru du 8.0101	31068	264	204.12	47	11
	P82952	γ-Conglutin 1	Pru du γ-conglutin	46945	431	202.36	33	10
AEP 6	E3SH28	Prunin 1	Pru du 6.0101	63052	551	513.70	82	259
	Q43607	Prunin 1	Pru du 6	63017	551	509.53	81	257
	Q43608	Prunin 2 (Fragment)	Pru du 6	57052	504	337.12	53	44
	E3SH29	Prunin 2 (Fragment)	Pru du 6, Pru du 6.0201	57006	504	334.88	56	45
	Q945K2	Mandelonitrile lyase 2	Pru du 10	61158	563	289.24	34	29
	A0A516F3L2	Cysteine-rich antimicrobial protein	Pru du 8, Pru du 8.0101	31068	264	214.49	50	15
	P82952	γ-Conglutin 1	Pru du γ-conglutin	46945	431	267.59	45	18
AEP 7	E3SH28	Prunin 1	Pru du 6.0101	63052	551	600.56	90	343
	Q43607	Prunin 1	Pru du 6	63017	551	591.88	88	339
	E3SH29	Prunin 2 (Fragment)	Pru du 6, Pru du 6.0201	57006	504	488.00	75	113
	Q43608	Prunin 2 (Fragment)	Pru du 6	57052	504	470.56	65	103
	A0A516F3L2	Cysteine-rich antimicrobial protein	Pru du 8, Pru du 8.0101	31068	264	219.98	45	12
	Q945K2	Mandelonitrile lyase 2	Pru du 10	61158	563	213.20	23	12
	P82952	γ-Conglutin 1	Pru du γ-conglutin	46945	431	108.38	8	3
AEP 8	E3SH28	Prunin 1	Pru du 6.0101	63052	551	506.61	87	243
	Q43607	Prunin 1	Pru du 6	63017	551	492.43	85	239
	E3SH29	Prunin 2 (Fragment)	Pru du 6, Pru du 6.0201	57006	504	488.41	72	231
	Q43608	Prunin 2 (Fragment)	Pru du 6	57052	504	439.33	66	194
	A0A516F3L2	Cysteine-rich antimicrobial protein	Pru du 8, Pru du 8.0101	31068	264	212.65	38	11
	Q945K2	Mandelonitrile lyase 2	Pru du 10	61158	563	181.35	25	10
	P82952	γ-Conglutin 1	Pru du γ-conglutin	46945	431	141.29	18	6
AEP 9	E3SH28	Prunin 1	Pru du 6.0101	63052	551	545.32	84	246
	Q43607	Prunin 1	Pru du 6	63017	551	536.21	82	237
	E3SH29	Prunin 2 (Fragment)	Pru du 6, Pru du 6.0201	57006	504	433.10	74	111
	Q43608	Prunin 2 (Fragment)	Pru du 6	57052	504	419.01	72	99
	A0A516F3L2	Cysteine-rich antimicrobial protein	Pru du 8, Pru du 8.0101	31068	264	395.81	74	54
	P82952	γ-Conglutin 1	Pru du γ-conglutin	46945	431	277.60	52	18
	Q945K2	Mandelonitrile lyase 2	Pru du 10	61158	563	277.39	43	26
	B6CQR7	PR-10	Pru du 1	17,636	160	182.58	40	6
	Q8GSL5	Profilin	Pru du 4, Pru du 4.0101,Pru du 4.0102	14061	131	56.15	10	1

EAEP 1	E3SH28	Prunin 1	Pru du 6.0101	63052	551	475.5	67	144
	Q43607	Prunin 1	Pru du 6	63017	551	461.55	69	145
	Q43608	Prunin 2 (Fragment)	Pru du 6	57052	504	316.32	38	36
	E3SH29	Prunin 2 (Fragment)	Pru du 6, Pru du 6.0201	57006	504	307.9	41	36
	Q945K2	Mandelonitrile lyase 2	Pru du 10	61158	563	263.89	31	18
	A0A516F3L2	Cysteine-rich antimicrobial protein	Pru du 8, Pru du 8.0101	31068	264	212.63	37	11
	P82952	γ-Conglutin 1	Pru du γ-conglutin	46945	431	190.75	24	8
EAEP 2	E3SH28	Prunin 1	Pru du 6.0101	63052	551	544.71	59	292
	Q43607	Prunin 1	Pru du 6	63017	551	538.75	60	295
	E3SH29	Prunin 2 (Fragment)	Pru du 6, Pru du 6.0201	57006	504	312.09	48	39
	Q43608	Prunin 2 (Fragment)	Pru du 6	57052	504	280.98	38	34
	A0A516F3L2	Cysteine-rich antimicrobial protein	Pru du 8, Pru du 8.0101	31068	264	110.47	13	3
	Q945K2	Mandelonitrile lyase 2	Pru du 10	61158	563	92.56	6	3
EAEP 3	E3SH28	Prunin 1	Pru du 6.0101	63052	551	546.9	74	237
	Q43607	Prunin 1	Pru du 6	63017	551	538.47	76	236
	E3SH29	Prunin 2 (Fragment)	Pru du 6, Pru du 6.0201	57006	504	521.84	62	220
	Q43608	Prunin 2 (Fragment)	Pru du 6	57052	504	492.24	62	195
	A0A516F3L2	Cysteine-rich antimicrobial protein	Pru du 8, Pru du 8.0101	31068	264	187.12	29	7
EAEP 4	Q43607	Prunin 1	Pru du 6	63017	551	568.8	83	299
	E3SH28	Prunin 1	Pru du 6.0101	63052	551	567.59	85	298
	E3SH29	Prunin 2 (Fragment)	Pru du 6, Pru du 6.0201	57006	504	449	65	87
	Q43608	Prunin 2 (Fragment)	Pru du 6	57052	504	431.8	59	80
	A0A516F3L2	Cysteine-rich antimicrobial protein	Pru du 8, Pru du 8.0101	31068	264	253.92	43	21
	Q945K2	Mandelonitrile lyase 2	Pru du 10	61158	563	150.79	14	7
	P82952	γ-Conglutin 1	Pru du γ-conglutin	46945	431	70.07	5	2
EAEP 5	E3SH28	Prunin 1	Pru du 6.0101	63052	551	613.12	83	387
	Q43607	Prunin 1	Pru du 6	63017	551	571.58	85	385
	E3SH29	Prunin 2 (Fragment)	Pru du 6, Pru du 6.0201	57006	504	434.04	69	139
	Q43608	Prunin 2 (Fragment)	Pru du 6	57052	504	409.93	65	118
	A0A516F3L2	Cysteine-rich antimicrobial protein	Pru du 8, Pru du 8.0101	31068	264	311.52	56	19
	Q945K2	Mandelonitrile lyase 2	Pru du 10	61158	563	123.33	8	4

	P82952	γ-Conglutin 1	Pru du γ-conglutin	46945	431	71.83	5	2

[1] Numbers of amion acids.

[2] −10lgP was determined by PEAKS Studio in database search.

8.3.1.3. Impact of extraction conditions on prunin 1 and prunin 2 epitope sequences

Proteolysis can reduce the molecular weight of proteins by breaking them down into smaller fragments. SDS-PAGE and proteomics analysis results for the gel cut samples revealed the reduction of several protein allergens after the use of neutral protease in the extraction. However, epitopes eliciting allergenic responses, especially linear epitopes, might still be intact after partial proteolysis. Therefore, further investigations on whether immunoreactive epitopes are affected by the protease are necessary.

Willison et al. (2011) identified six and eight IgE-reactive epitope sequences from prunin 1 (Pru du 6.01) and prunin 2 (Pru du 6.02), respectively, using overlapping peptides and a pooled serum from almond allergic patients. To better understand the impact of using neutral protease during extraction on the epitope sequences in almond proteins, the relative abundance of these sequences was estimated by calculating the proportion of the epitope sequence regions among all the identified tryptic peptides (Table 8.2). The results showed that the epitope sequence of prunin 1 f(161–175) represented a significant abundance in the gel cuts AEP 1 and AEP 3–6 (47‰ – 133‰). The abundance of the same epitope sequence in the EAEP gel cuts was much lower than in the AEP ones, demonstrating that the enzyme greatly destroyed this epitope at the protein level. Overall, lower epitope abundances in the EAEP samples were observed for several epitope sequences, including prunin 1 f(145–159), prunin 2 f(185–199), prunin 2 f(209–223), and prunin 2 f(281–295) compared with AEP samples. These epitope sequences are all located in the α-subunit region, which suffered extensive proteolysis during EAEP (Fig. 8.1C).

In contrast, the epitope sequences in the β-subunit region, including prunin 1 f(510–524) and prunin 2 f(465–479), generally had similar or higher relative abundance in the EAEP gel samples compared to the AEP. AEP 7 and EAEP 2, which were the main bands containing prunin 1 β-subunit polypeptide, had comparable relative abundances of prunin 1 f(510–524) (38‰ and 31‰, respectively). Similarly, AEP 8 and EAEP 3, which consisted of mainly prunin 2 β-subunit polypeptide, contained similar relative abundances of prunin 2 f(465–479) (16‰ and 14‰, respectively). This reveals that a significant portion of the two epitope sequences was still encrypted in the β-subunit polypeptide despite the use of the enzyme in EAEP. The resistance of the epitope sequences of prunin 1 f(510–524) and prunin 2 f(465–479), located in the β-subunit regions, is in agreement with the resistance of the β-subunits against hydrolysis by the neutral protease. Because the protease substantially reduced the α-subunits during EAEP, the β-subunits became dominant in all the EAEP gel samples (Fig. 8.1C). This could, in turn, explain the higher relative abundance of prunin 1 f(510–524) and prunin 2 f(465–479) from β-subunits in EAEP 1, EAEP 4, and EAEP 5 (Table 8.2).

Besides being encrypted in proteins and larger peptides that can be detected by Tris-glycine SDS-PAGE (~6–400 kDa), linear epitopes may also exist in low-molecular-weight peptides (<5 kDa). Low-molecular-weight peptides can be naturally occurring or generated by enzymatic hydrolysis. Our previous study characterized low-molecular-weight peptides present in the almond skim fraction (protein-rich extract) generated by AEP and EAEP with LC-MS/MS by searching against the UniProt protein database (Huang et al., 2022). In that study, a total of 523 and 1009 low-molecular-weight peptides were identified from AEP and EAEP skims, respectively, with the majority originating from prunin 1 (193 and 96 sequences, respectively) and prunin 2 (194 and 133 sequences, respectively). By searching the IgE-reactive epitope sequences reported by

Willison et al., 2011, 25 peptide sequences in the AEP extract were found to contain the full sequence of any of the IgE-reactive epitopes (Table 8.3). The IgE-reactive epitope sequences involved in the 25 peptides include prunin 1 f(161–175) (seven peptides), prunin 1 f(225–239) (seven peptides), prunin 1 f(510–524) (nine peptides), prunin 2 f(121–135) (one peptide), and prunin 2 f(209–223) (one peptide). Conversely, none of the peptides present in the EAEP samples contain any IgE-reactive epitopes in a full sequence, demonstrating that proteolysis disrupted IgE-reactive epitopes encrypted in low-molecular-weight peptides.

Table 8.2. Relative abundance (‰)[a] of epitope sequences in prunin 1 and prunin 2 in SDS-PAGE gel spot samples

Epitope sequence[b]	AEP 1	AEP 2	AEP 3	AEP 4	AEP 5	AEP 6	AEP 7	AEP 8	AEP 9	EAEP 1	EAEP 2	EAEP 3	EAEP 4	EAEP 5
Prunin 1 α-subunit														
118 SSQQGRQQEQEQERQ 132	4.534[c]	0.518	8.611	5.644	10.425	1.262	0.451	1.005	4.392	0.035	0	0.003	4.934	1.528
145 QQEQQQERQGRQQGR 159	7.609	1.184	14.493	10.907	19.739	2.165	1.732	4.499	1.974	0.071	0	0	0.057	0
161 QQEEGRQQEQQQGQQ 175	47.034	14.084	104.519	98.924	132.942	57.731	4.896	3.212	11.409	0.546	0.006	0.011	0.066	0.027
225 LFHVSSDHNQLDQNP 239	2.470	0.630	7.652	7.852	5.544	5.579	0.396	0.497	4.632	0.389	0.004	0.011	1.447	7.381
281 QQEQQGSGNNVFSGF 295	0.020	0.000	0.147	0.017	0.036	0.073	0.005	0.013	0.066	0.000	0	0.000	0.001	0.013
Prunin 1 β-subunit														
510 RALPDEVLANAYQIS 524	5.260	0.858	1.340	1.177	2.804	3.658	38.079	5.066	2.820	13.485	31.368	18.367	18.931	25.413
Prunin 2 α-subunit														
17 FGQNKEWQLNQLEAR 31	0	0	0	0	0	0	0	0	0	0	0	0	0	0
105 DSQPQQFQQQQQQQQ 119	0.165	0.011	0.000	0.147	0.042	0.014	0.026	0	0.090	0	0	0	0.038	0
121 RPSRQEGGQGQQQFQ 135	0.245	0.006	0.033	0.419	0.160	0.100	2.627	0.424	1.428	0.031	0.001	0.036	0.358	1.088
185 QNQLDQVPRRFYLAG 199	1.071	0.100	0.074	3.116	0.523	0.219	0.733	0.094	1.159	0.017	0.002	0.026	0.047	0.157
209 QQGRQQQQQQQGQQG 223	0.874	0.078	0.053	2.386	0.369	0.169	0.500	0.073	0.949	0.013	0.002	0.013	0.035	0.067
225 GNNIFSGFDTQLLAQ 239	0.004	0.000	0	0.002	0.002	0.000	0.001	0.000	0.030	0	0	0	0	0
281 RGDQERQQEEQQSQR 295	0.545	0.123	0.032	0.335	0.027	0.153	0.129	0.021	0.050	0	0	0	0.001	0
Prunin 2 β-subunit														
465 QNAFRISRQEARNLK 479	0.545	0.200	0.581	1.156	0.174	0.425	0.299	15.808	3.679	1.166	0.213	14.189	1.383	6.075

High to low

[a] Relative abundance was expressed in the total peak area of the epitope region/total peak area of the identified peptides in the same gel spot.

[b] Epitope sequences were reported by Willison et al., 2011.

[c] Background color from red to green represents abundance from high to low.

Table 8.3. Low-molecular-weight peptides[a] containing epitope sequences from prunin 1 and prunin 2 found in the AEP and EAEP skims

Epitope sequence[b]	Peptide sequence	
	AEP	EAEP
Prunin 1 α-subunit		
118 SSQQGRQQEQEQERQ 132	-	-
145 QQEQQQERQGRQQGR 159	-	-
161 QQEEGRQQEQQQGQQ 175	GRQQQEEGRQQEQQQGQQGRPQ, RQQQEEGRQQEQQQGQQGRPQQ, GRQQQEEGRQQEQQQGQQGRPQQ, GRQQQEEGRQQEQQQGQQGRPQQQ, GRQQQEEGRQQEQQQGQQGRPQQQQ, GRQQQEEGRQQEQQQGQQGRPQQQQQ, GRQQQEEGRQQEQQQGQQGRPQQQQQFRQ	-
225 LFHVSSDHNQLDQNP 239	LFHVSSDHNQLDQNPRK, YNDGDQELVAVNLFHVSSDHNQLDQNPRK, WSYNDGDQELVAVNLFHVSSDHNQLDQNPRK, YNDGDQELVAVNLFHVSSDHNQLDQNPRKFY, YWSYNDGDQELVAVNLFHVSSDHNQLDQNPRK, VAYWSYNDGDQELVAVNLFHVSSDHNQLDQNPRK, VAYWSYNDGDQELVAVNLFHVSSDHNQLDQNPRKF	-
281 QQEQQGSGNNVFSGF 295	-	-
Prunin 1 β-subunit		-
510 RALPDEVLANAYQIS 524	LRALPDEVLANAYQISREQ, FLRALPDEVLANAYQISREQ, LRALPDEVLANAYQISREQARQ, FLRALPDEVLANAYQISREQARQ, LRALPDEVLANAYQISREQARQL, LRALPDEVLANAYQISREQARQLK, FLRALPDEVLANAYQISREQARQLK, LRALPDEVLANAYQISREQARQLKY, FLRALPDEVLANAYQISREQARQLKY	-
Prunin 2 α-subunit		
17 FGQNKEWQLNQLEAR 31	-	-
105 DSQPQQFQQQQQQQ 119	-	-
121 RPSRQEGGQGQQQFQ 135	FRPSRQEGGQGQQQFQGEDQQDRHQK	-
185 QNQLDQVPRRFYLAG 199	-	-
209 QQGRQQQQQQQGQQG 223	LAGNPQDEFNPQQQGRQQQQQQQGQQGNGNNIFSGFDTQ	-
225 GNNIFSGFDTQLLAQ 239	-	-
281 RGDQERQQEEQQSQR 295	-	-
Prunin 2 β-subunit		
465 QNAFRISRQEARNLK 479	-	-

[a] Low molecular weight peptides were identified from AEP and EAEP extracts by Willison et al., 2011.
[b] Epitope sequences were reported by Willison et al., 2011.

8.3.2. Effects of protein hydrolysis on in vitro protein digestibility

Extraction conditions, especially proteolysis, can significantly affect the protein *in vitro* digestibility of the extracted protein. Protein hydrolysates have been reported as a suitable source of protein for human nutrition, as their gastrointestinal absorption seems to be more effective than that of intact proteins (Grimble, 1991).

The *in vitro* digestibility of almond proteins from the unhydrolyzed (AEP) and hydrolyzed samples (EAEP) are shown in Fig. 8.2A. Casein and soy protein isolated powder were also evaluated for comparison purposes. Proteolysis significantly improved the *in vitro* protein digestibility of almond protein samples from 79.1 ± 2.4% to 88.5 ± 3.6%. While casein and soy protein exhibited the highest and lowest digestibility values (92.9 ± 2.7% and 72.3 ± 4.3%, respectively, casein digestibility was not significantly different from that of the almond hydrolysates (EAEP). Casein is commonly used as a reference protein for *in vitro* protein digestibility assays, and the value herein reported agrees with the ones reported for casein digestibility (92-99%) (Alonso et al., 2000; El-Aal et al., 1986). The higher protein digestibility of the EAEP samples can be attributed to the partial breakdown (moderate degree of protein hydrolysis) of large protein bodies into smaller sizes by the protease (He et al., 2015), which can facilitate the access of digestive enzymes (pepsin and pancreatin) to the protein sites due to reduction in steric hindrance, leading to an improvement in the protein hydrolysis during digestion.

Similar findings have been shown for air-classified pea protein-enriched flour where protein digestibility increased from 84 to 89% after hydrolysis by papain (DH of 11%) (Konieczny et al., 2020). Our results differ from the ones reported by de Souza, Dias, Oliveira, de Moura Bell, & Koblitz, 2020 in that a decrease in the *in vitro* protein digestibility from 73 to 64% was observed after the use of alkaline protease to assist the extraction of protein and oil from

the almond cake. It is important to highlight that besides the difference in starting material (full-fat almond flour in our study vs. almond cake in theirs) the aforementioned study used a different protease (alkaline protease) to assist the extraction and the extracted protein had a significantly higher degree of hydrolysis (7 in ours vs 23% in theirs) before digestion. Extensive hydrolysis could entail fewer attack sites available for the digestive enzymes (pepsin and pancreatin), which could have underestimated this parameter (Souza, Dias, Oliveira, de Moura Bell, & Koblitz, 2020).

The protein profile of AEP and EAEP samples during oral (0.5 min), gastric (30, 60, 90, 120 min), and duodenal (150, 180, 210, and 240 min) digestion was evaluated by SDS-PAGE (Fig. 8.2B, C). The oral phase did not affect the protein profile of AEP and EAEP samples as expected. However, after 30 min of digestion, significant proteolysis was observed for both samples, with AEP and EAEP samples having the majority of protein bands < 18 kDa and < 14 kDa, respectively. Our results agree with the ones reported by Sathe, 1993, where only peptides < 20 kDa were found after 30 min of gastric digestion of the almond flour protein, and demonstrate that proteolysis before digestion can enhance protein hydrolysis by gastric enzymes. Similar results were reported by Souza et al. 2020, where faster digestion of almond cake hydrolysates by pepsin was observed during *in vitro* digestion.

From 60 to 90 min of gastric digestion the protein profile remained the same. At 120 min, the AEP protein profile was slightly more hydrolyzed than after 90 min. After 30 min of duodenal digestion (total digestion time of 150 min), only a small band at ~10 kDa can be seen for both AEP and EAEP samples, indicating that all proteins were broken down into protein fragments and peptides < 10 kDa. This band gets fainter with the increase of duodenal digestion time, indicating that the protein sites susceptible to proteolysis are accessible to the digestive enzymes (pepsin and pancreatin) in almond proteins. Similar results were reported by Souza et al, 2020 for almond cake

proteins where no bands were observed in the SDS-PAGE after intestinal digestion. It is important to mention that the bands found in the gastric phase for both AEP and EAEP samples at ~38 kDa and the bands found in the intestinal phase from 25–55 kDa are related to the pepsin and pancreatin enzymes, respectively (Fig. 8.2 B, C and Supplementary material Fig. 8.S3).

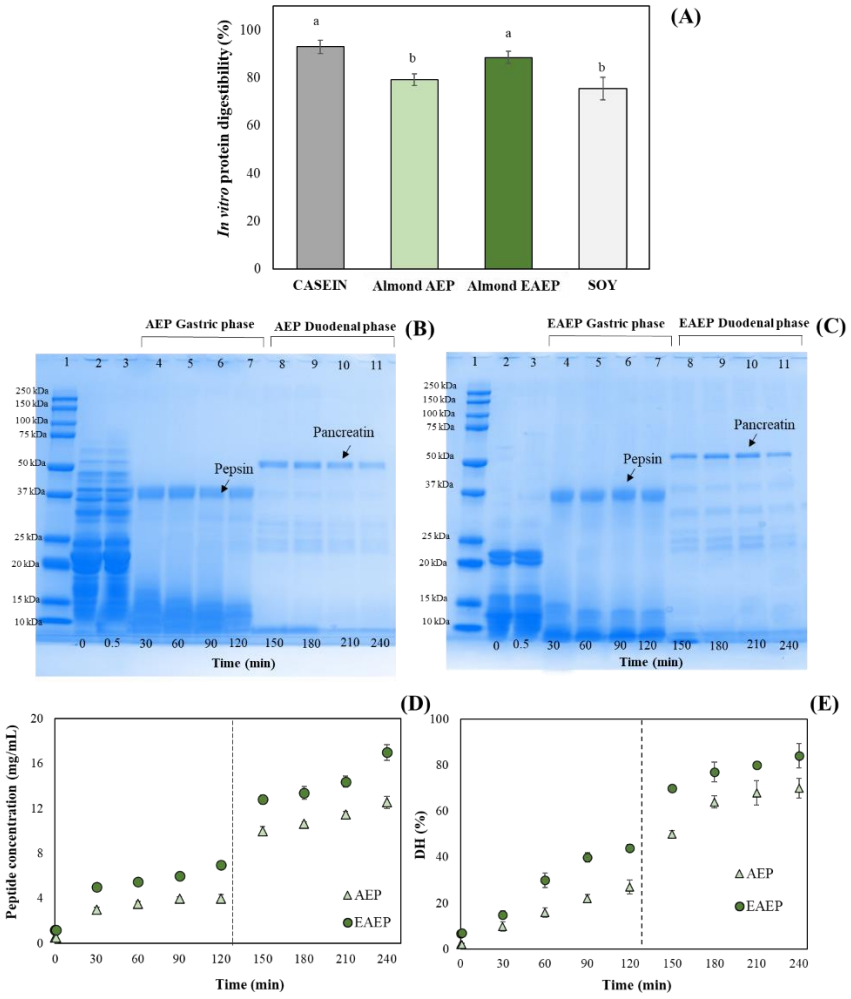


Fig. 8.2. Total *in vitro* protein digestibility of casein, unhydrolyzed (AEP) and hydrolyzed (EAEP) almond protein extracts and soybean isolated protein. Different letters indicate a significant difference between samples at $p < 0.05$ (A). SDS-PAGE of digestion kinetics for the AEP (B) and EAEP (C) almond proteins samples, arrows indicate the pepsin in the gastric phase and the pancreatin in the intestinal phase. Peptide concentration (D) and degree of hydrolysis (DH) (E) of AEP and EAEP proteins samples.

The peptide quantification and the degree of hydrolysis (DH) of AEP and EAEP samples during digestion were also accessed at 0, 0.5, 30, 60, 90, 120, 150, 180, 210, and 240 min (Fig. 8.2 B and C). The hydrolyzed sample (EAEP) presented a higher peptide content (1.7 vs. 2.9 mg/mL) and a higher DH (1.5 vs. 7%) before digestion due to the action of the protease during the extraction process compared with the unhydrolyzed sample (AEP). Oral digestion did not affect the peptide content or the DH of both AEP and EAEP samples. During gastric digestion, a steady increase in the peptide concentration and DH was observed due to the action of pepsin. EAEP samples presented significantly higher peptide concentration and DH values compared with the unhydrolyzed samples (AEP) during the gastric phase. At the end of the gastric phase, the DH of the AEP and EAEP samples reached 21 and 41%, respectively. During the duodenal digestion, a further increase in the peptide concentration and DH was observed due to the action of pancreatic enzymes. While AEP and EAEP samples showed similar trends, EAEP samples were more extensively hydrolyzed through the course of duodenal digestion. The significant increase in the peptide concentration and DH observed after 30 min of intestinal digestion (150 min of total digestion time) can be attributed to the pepsin action during the gastric phase, which promoted the hydrolysis of the almond protein thus facilitating access to the protein sites for the pancreatin enzyme. The peptide concentration and the DH significantly increased within digestion time reaching values of 17 and 13 mg/mL for peptide concentration and 86 and 71% DH for EAEP and AEP samples, respectively. The peptide concentration and DH are in accordance with the SDS-PAGE protein profile observed for AEP and EAEP proteins. The higher peptide concentration and DH of the EAEP samples suggest a greater exposure of cleavage sites in the smaller protein fragments present in this sample. Those results are in congruence with the higher total *in vitro* digestibility of the almond hydrolyzed (EAEP) samples and reinforce the beneficial role of using

selected proteases to assist the extraction of full-fat almond flour as an effective strategy to significantly enhance protein *in vitro* digestibility.

8.3.3. Effect of enzymatic hydrolysis on protein antigenicity

8.3.3.1. Sandwich ELISA for almond immunoreactivity

Protein hydrolysis has been used in the production of hypoallergenic food ingredients because of its effectiveness in disrupting sequential and conformational epitopes (Cabanillas et al., 2012; Verhoeckx et al., 2015). However, depending on the type of enzymes used and the hydrolysis conditions, peptides of different lengths may be obtained carrying more or less allergenicity (Cabanillas et al., 2010, Cabanillas et al., 2012; Clemente, Vioque, Sanchez-Vioque, Pedroche, & Millán, 1999)

Aiming to understand the impact of enzymatic extraction on almond protein immunoreactivity, a preliminary assessment of the potential allergenicity of the almond protein extract was performed using a rabbit antibody-based inhibition ELISA assay to detect and quantify the presence of major almond allergenic protein (amandin). Due to the lack of manufacturer's information about the almond allergen which the antibody is raised against, the levels of immunoreactivity recorded for the almond protein sample were considered representative of the total allergenicity. Fig. 8.3A illustrates the ELISA results for AEP and EAEP almond protein samples. Enzyme hydrolysis promoted a 75% reduction in the total immunoreactivity of almond proteins estimated by ELISA. These results suggest that the use of enzyme during the extraction affected the structural conformation of almond proteins in a manner that reduced the detection of the almond protein by the assay. These preliminary results indicate a potential reduction in the allergenicity of the almond hydrolysates. Similar results were reported by Clemente et al., 1999 for chickpea protein where an 80% antigenicity reduction was achieved after hydrolysis with

Alcalase using an antibody-capture assay. Proteolysis was also reported as an efficient strategy to reduce the antigenicity of lentil and peanut proteins accessed using ELISA test and serum pool from patients with clinical allergy to lentil and peanut (Cabanillas et al., 2010, 2012).

8.3.3.2. Immunoreactivity of AEP and EAEP protein extracts by Western blotting

In order to better understand the observed effects of proteolysis on the almond protein antigenicity, Western immunoblotting (IgE and IgG) using human sera was performed using human blood serum from two previously selected patients (4C and 196b) showing reactivity to almonds (Fig. 8.3C, D). The IgE immunoblot assay showed recognition of Prunin (60 kDa Pru du 6), Prunin α-subunit (40 kDa), and Prunin 1 and 2 β-subunits (21 kDa and 19 kDa) for the AEP samples (lanes A1-A3) for 4C and 196b human sera with the more intense antigenic response being observed for human serum sample 196b (Fig. 8.3C). Prunins are constitutes of amandin, the almond major protein. Roux, Teuber, Robotham, & Sathe, 2001 reported that amandin is an excellent marker protein for protein allergenicity since it accounts for the majority of total almond protein and is substantially heat-stable. Although other proteins may also be implicated in almond food allergies for a particular patient, amandin appears to include the key IgE-reactive polypeptides in sera from patients with life-threatening almond food allergy (Roux, Teuber, Robotham, & Sathe, 2001).

Proteolysis significantly reduced IgE recognition in both human sera (Fig. 8.3B, C). Compared with the unhydrolyzed samples (AEP), there was no recognition of proteins above 22 kDa, similar recognition of proteins at ~20 kDa, and a reduction in the recognition at 19 kDa in the hydrolysates (EAEP). However, smaller protein fragments in the EAEP samples were also recognized. Human serum sample 196b showed more intense bands in comparison with human serum sample 4C for EAEP (lanes B1-B3) proteins. Proteins at ~20 kDa are probably more

resistant to hydrolysis than other immunoreactive proteins as observed in the SDS-PGAE gel and proteomics assays (Fig. 8.1). Therefore, proteolysis resulted in important destruction of IgE-binding epitopes in the almond hydrolysates as shown by *in vitro* experiments. However, some allergenic proteins were still detected by sera from the two tested patients. Significant reduction in IgE reactivity was also reported for lentils (Cabanillas et al., 2010a) and peanut (Cabanillas et al., 2012) protein hydrolysates using immunoblotting and ELISA test.

The IgG immunoblot assay (Fig. 8.3D, E) also showed recognition of similar proteins as the IgE assay for both AEP and EAEP samples, however, more bands above 40 kDa were reactive for the unhydrolyzed samples. Protein hydrolysis also promoted a reduction in IgG recognition of almond proteins for both 4C and 196b human sera samples.

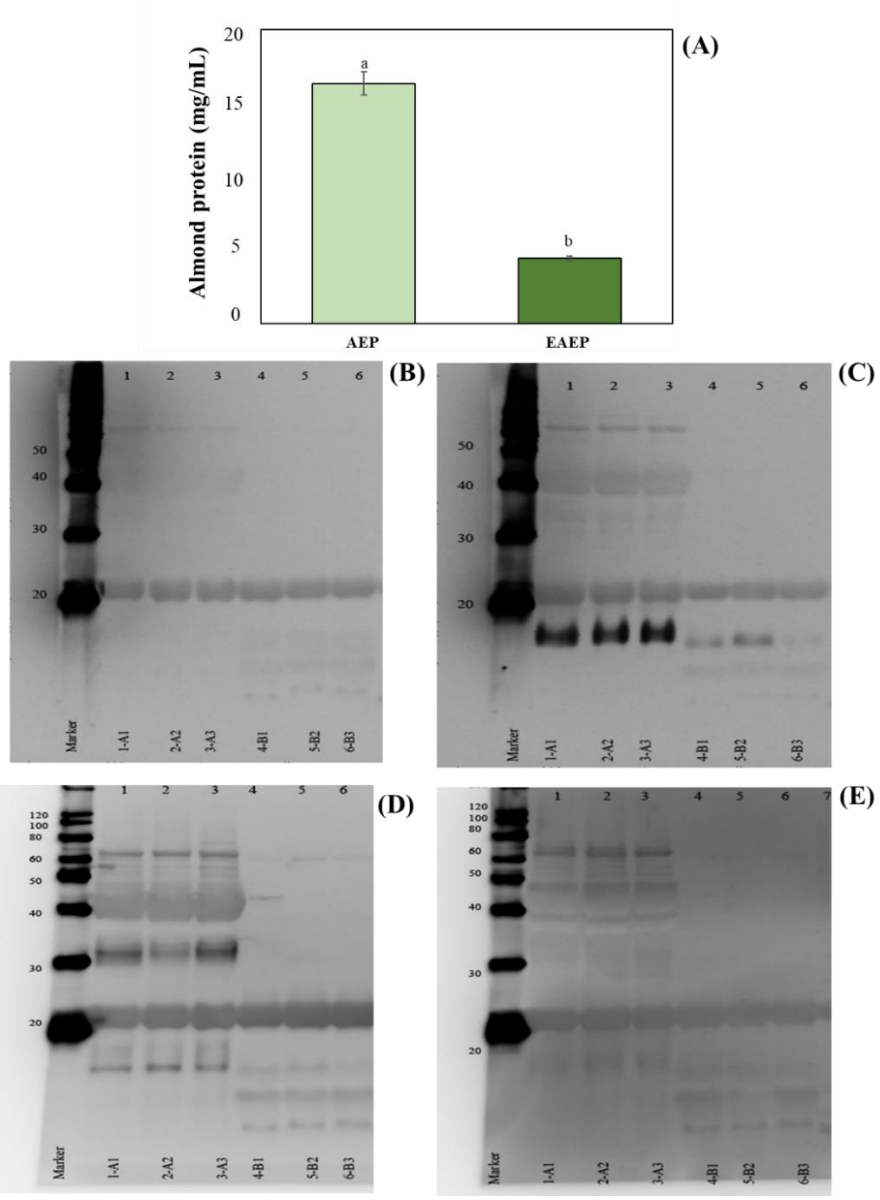


Fig. 8.3. Total almond protein immunoreactivity by sandwich ELISA (A). Different letters indicate a significant difference between samples at $p < 0.05$. Western blot of AEP (unhydrolyzed samples - A1 to A3) and EAEP (B1 to B3): Human serum sample 4C IgE, Primary: human sera 1:20, Secondary: mouse anti-Human IgE Fe 1:5000 (B); Human serum sample 196b IgE, Primary: human sera 1:20, Secondary: mouse anti-Human IgE Fe 1:5000 (C); Human serum sample 4C IgE, Primary: human sera 1:200, Secondary: goat anti-Human IgG Fe 1:10,000 (D); Human serum sample 196b IgE, Primary: human sera 1:200, Secondary: goat anti-Human IgG Fe 1:10,000 (E).

Due to its high reactivity, IgE immunoblot using 196b human serum sample was selected to further investigate the reactivity of the proteins at 18-19 kDa, which showed higher immunoreactivity for both AEP and EAEP samples in the western blot assay (Fig. 4B–E). The protein bands were subjected to integration using Image J software (Fig. 8.4A, C). A 74% reduction in the area value was observed for the hydrolyzed samples, in accordance with the Elisa results (Fig. 8.4B). Overall, the results indicate that the use of a neutral protease to assist the extraction of almond proteins resulted in structural protein changes that decreased IgE and IgG recognition compared to the unhydrolyzed samples. Those results are in accordance with the proteomics results that reported partial destruction of the allergen protein epitopes and a reduction in its detection in the EAEP samples.

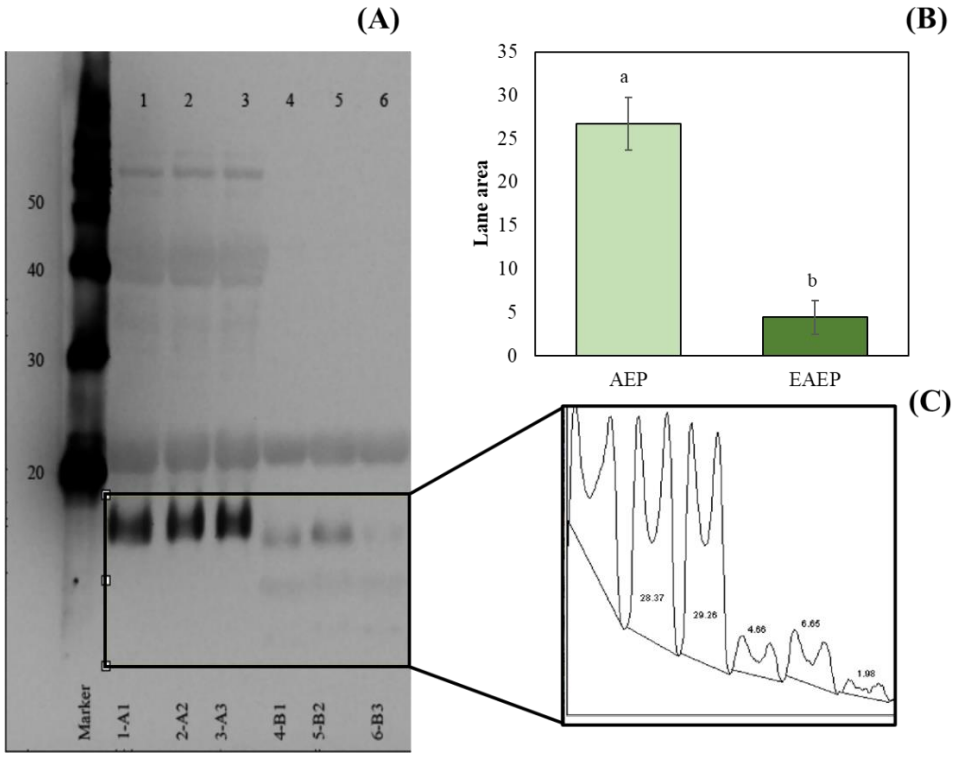


Fig. 8.4. Western blot of 196b IgE highlighting the more reactive bands (A); Average of the lane area for the AEP and EAEP samples (B), Integration of the highlighted bands by Image J (C).

8.4. Conclusions

This study sheds light on the impact of the use of a neutral protease to assist the extraction of proteins from full-fat almond flour on the digestibility and allergenicity of the extracted proteins. The use of protease during the extraction process (EAEP) led to the total hydrolysis of prunin 1 and 2 α-subunit and partial hydrolysis of prunin 1 β-subunit. Proteolysis led to the formation of smaller protein fragments and peptides and a reduction in the protein allergen epitopes identification. Importantly, protein hydrolysis also significantly improved the protein *in vitro* digestibility from 79.1 to 88.5%, as evidenced by the higher release of peptides and degree of hydrolysis during the gastric and duodenal digestion phases. A reduction in 74% of immunogenicity was observed for the hydrolyzed samples along with a reduction in the IgE and IgG recognition compared to the unhydrolyzed almond proteins. Enzymatic extraction of almond proteins led to the production of protein hydrolases with improved digestibility and reduced antigenicity that could be an alternative to the use of intact protein in the development of hypoallergenic food ingredients. Although further studies are needed to characterize the biological activity of the residual allergens and to assess the clinical relevance of our findings, this enzymatic procedure seems to be a promising method to obtain hypoallergenic protein hydrolysates.

CRediT author contribution statement

Fernanda F. G. Dias: Methodology, Data curation, Formal analysis, Writing- Original draft preparation, Writing- Reviewing and Editing. **Yu-Ping Huang:** Methodology, Data curation, Formal analysis, Writing- Reviewing and Editing. **Joseph Schauer**: Methodology, Formal analysis, Writing- Reviewing and Editing**. Daniela Barile:** Supervision. Reviewing and Editing.

Judy Van de Water: Supervision. Reviewing and Editing. **Juliana M. L. N. de Moura Bell:** Supervision, Project administration; Resources; Writing, Reviewing and Editing.

Acknowledgments

We would like to acknowledge the U.S. Department of Agriculture's (USDA) Agricultural Marketing Service through grant AM170100XXXX G011. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the USDA.

Conflict of interest

The authors declare that they have no conflict of interest.

Supplementary materials

Table 8.S1. Composition of the simulated fluids.

Simulated Saliva Fluid (SSF)	Final concentration (mg/mL)	Final pH
Amylase	2.5 (75 U/mL)	
Mucin	1.0	
NaCl	0.117	7.0
KCl	0.149	
$NaHCO_3$	2.1	
Simulated Gastric Fluid (SGF)	Final concentration (mg/mL)	Final pH
Pepsin	0.75 (2000 U/mL)	
Gastric mucin	1.5	1.8–2.0
NaCl	8.78	
Simulated Intestinal Fluid (SIF)	Final concentration (mg/mL)	Final pH
Pancreatin	8.0 (800 U/mL)	
Bile extract	10	7.0
$NaHCO_3$	16.8	

Table 8.S2. Identification of AEP1 band from SDS-PAGE by LC-MS/MS analysis of tryptic peptides produced by in-gel protein digestion.

Protein accession	Protein name	Coverage	Area	Protein accession for coverage and area	Relative abundance (%)
sp|E3SH28|PRU01:tr|A0A5E4FFS0|A0A5E4FFS0:sp|Q43607|PRU1	Prunin 1	83	4.20E+10	sp|E3SH28|PRU01	66.0
tr|Q43608|Q43608:tr|E3SH29|E3SH29:tr|A0A5E4FK23|A0A5E4FK23:tr|A0A4Y1S2I9|A0A4Y1S2I9	Prunin 2	62	6.34E+09	tr|E3SH29|E3SH29	10.0
sp|Q945K2|MDL2	(R)-mandelonitrile lyase 2	50	5.85E+09	sp|Q945K2|MDL2	9.2
tr|A0A5E4EZP4|A0A5E4EZP4:tr|A0A4Y1QPK2|A0A4Y1QPK2	PREDICTED: vicilin	55	3.96E+09	tr|A0A5E4EZP4|A0A5E4EZP4	6.2
tr|A0A5E4E244|A0A5E4E244:tr|A0A4Y1RVW6|A0A4Y1RVW6	PREDICTED: vicilin	37	1.37E+09	tr|A0A5E4E244|A0A5E4E244	2.1
tr|A0A5E4EE27|A0A5E4EE27	PREDICTED: vicilin	37	9.15E+08	tr|A0A5E4EE27|A0A5E4EE27	1.4
tr|A0A5E4ED32|A0A5E4ED32	Protein disulfide-isomerase	44	4.49E+08	tr|A0A5E4ED32|A0A5E4ED32	0.7
tr|A0A4Y1QPI1|A0A4Y1QPI1:tr|A0A5E4F2T7|A0A5E4F2T7:tr|A0A5J6V1A4|A0A5J6V1A4	PREDICTED: vicilin	30	4.24E+08	tr|A0A4Y1QPI1|A0A4Y1QPI1	0.7
tr|A0A5E4EYX0|A0A5E4EYX0:tr|A0A5E4EYT9|A0A5E4EYT9:tr|A0A516F3L2|A0A516F3L2	Cysteine rich antimicrobial protein	52	2.15E+08	tr|A0A5E4EYX0|A0A5E4EYX0	0.3
tr|A0A5E4E767|A0A5E4E767:tr|A0A5E4E5K6|A0A5E4E5K6	PREDICTED: tripeptidyl-peptidase 2	26	1.45E+08	tr|A0A5E4E767|A0A5E4E767	0.2
tr|H9ZGE3|H9ZGE3:tr|H9ZGD9|H9ZGD9:tr|H9ZGE2|H9ZGE2:tr|H9ZGD8|H9ZGD8:tr|A0A5H2XK07|A0A5H2XK07	Prunasin hydrolase	27	1.32E+08	tr|H9ZGE3|H9ZGE3	0.2
tr|A0A5E4G2Q4|A0A5E4G2Q4	PREDICTED: seed	43	1.19E+08	tr|A0A5E4G2Q4|A0A5E4G2Q4	0.2
tr|A0A4Y1R8N9|A0A4Y1R8N9	Alpha-mannosidase	15	1.18E+08	tr|A0A4Y1R8N9|A0A4Y1R8N9	0.2
tr|A0A4Y1RHV0|A0A4Y1RHV0:tr|H9ZGE1|H9ZGE1:tr|A0A5E4FW87|A0A5E4FW87:tr|A0A4Y1RIL4|A0A4Y1RIL4	Beta glucosidase 17	22	1.18E+08	tr|A0A4Y1RHV0|A0A4Y1RHV0	0.2
tr|A0A5E4F4B3|A0A5E4F4B3:tr|A0A5E4GGU0|A0A5E4GGU0:tr|A0A5E4E9G9|A0A5E4E9G9:tr|A0A5E4F626|A0A5E4F626:tr|A0A4Y1RP48|A0A4Y1RP48:tr|A0A5E4F528|A0A5E4F528:tr|A0A5E4EFQ5|A0A5E4EFQ5:tr|A0A5E4FI37|A0A5E4FI37:tr|A0A4Y1RHX5|A0A4Y1RHX5	PREDICTED: heat shock	40	1.17E+08	tr|A0A5E4F4B3|A0A5E4F4B3	0.2
tr|A0A5E4ECQ1|A0A5E4ECQ1:tr|A0A5H2Y344|A0A5H2Y344	NAD(P)-binding Rossmann-fold superfamily protein	44	9.21E+07	tr|A0A5E4ECQ1|A0A5E4ECQ1	0.1
tr|A0A5E4FV72|A0A5E4FV72	PREDICTED: vicilin	37	8.37E+07	tr|A0A5E4FV72|A0A5E4FV72	0.1
tr|A0A5E4E543|A0A5E4E543	Malic enzyme	33	8.12E+07	tr|A0A5E4E543|A0A5E4E543	0.1

tr|A0A5E4EAH6|A0A5E4EAH6	PREDICTED: basic 7S globulin	40	7.11E+07	tr|A0A5E4EAH6|A0A5E4EAH6	0.1
tr|A0A5E4F1F2|A0A5E4F1F2:tr|A0A5E4F561|A0A5E4F561:tr|A0A4Y1RIB8|A0A4Y1RIB8	Malate synthase	33	6.69E+07	tr|A0A5E4F1F2|A0A5E4F1F2	0.1
sp|Q43804|OLEO1:tr|A0A5E4EAT1|A0A5E4EAT1	Oleosin 1	35	6.01E+07	sp|Q43804|OLEO1	0.1
tr|A0A5E4E8M3|A0A5E4E8M3:tr|A0A5E4E7H8|A0A5E4E7H8	PREDICTED: 17	17	4.34E+07	tr|A0A5E4E8M3|A0A5E4E8M3	0.1
tr|A0A5E4FFP9|A0A5E4FFP9:tr|A0A4Y1RB45|A0A4Y1RB45	Succinate-semialdehyde dehydrogenase	36	4.30E+07	tr|A0A5E4FFP9|A0A5E4FFP9	0.1
tr|A0A4Y1R8R9|A0A4Y1R8R9	Glycosyl hydrolase family 38 protein	15	4.30E+07	tr|A0A4Y1R8R9|A0A4Y1R8R9	0.1
tr|A0A4Y1RH78|A0A4Y1RH78:tr|A0A5E4F1W6|A0A5E4F1W6	Hydroxysteroid dehydrogenase 1	23	3.92E+07	tr|A0A4Y1RH78|A0A4Y1RH78	0.1
tr|A0A5E4EUA6|A0A5E4EUA6	Alpha-mannosidase	12	3.63E+07	tr|A0A5E4EUA6|A0A5E4EUA6	0.1
tr|A0A5E4ET55|A0A5E4ET55	Oleosin	30	3.49E+07	tr|A0A5E4ET55|A0A5E4ET55	0.1
tr|A0A5E4GE42|A0A5E4GE42:tr|A0A4Y1QME7|A0A4Y1QME7	Uncharacterized protein	23	3.49E+07	tr|A0A5E4GE42|A0A5E4GE42	0.1
tr|A0A4Y1QQ02|A0A4Y1QQ02:tr|A0A5E4F2W9|A0A5E4F2W9	Glyceraldehyde-3-phosphate dehydrogenase	27	3.10E+07	tr|A0A4Y1QQ02|A0A4Y1QQ02	0.0
tr|A0A5E4FBP4|A0A5E4FBP4	PREDICTED: non-classical arabinogalactan	23	3.09E+07	tr|A0A5E4FBP4|A0A5E4FBP4	0.0
tr|A0A5E4ER27|A0A5E4ER27:tr|A0A4Y1RJK8|A0A4Y1RJK8	PREDICTED: aldehyde dehydrogenase	25	2.98E+07	tr|A0A5E4ER27|A0A5E4ER27	0.0
tr|A0A5E4EH33|A0A5E4EH33:tr|F6K732|F6K732_9ROSA:tr|A0A5H2YB58|A0A5H2YB58	Pyruvate decarboxylase	14	2.39E+07	tr|A0A5E4EH33|A0A5E4EH33	0.0
tr|A0A5E4F5S2|A0A5E4F5S2:tr|A0A5E4F5X4|A0A5E4F5X4	Amine oxidase	15	2.31E+07	tr|A0A5E4F5S2|A0A5E4F5S2	0.0
tr|A0A5E4FTY1|A0A5E4FTY1	Annexin	15	2.22E+07	tr|A0A5E4FTY1|A0A5E4FTY1	0.0
tr|A0A5E4FAS8|A0A5E4FAS8	PREDICTED: fasciclin	14	2.07E+07	tr|A0A5E4FAS8|A0A5E4FAS8	0.0
tr|A0A5E4FUY2|A0A5E4FUY2:tr|A0A5E4FLL1|A0A5E4FLL1	Elongation factor 1-alpha	23	2.02E+07	tr|A0A5E4FUY2|A0A5E4FUY2	0.0
tr|A0A5E4G6M4|A0A5E4G6M4:tr|A0A5H2XK06|A0A5H2XK06	Late embryogenesis abundant protein family protein	27	1.89E+07	tr|A0A5E4G6M4|A0A5E4G6M4	0.0
tr|A0A5E4EFQ5|A0A5E4EFQ5	PREDICTED: luminal-binding	20	1.76E+07	tr|A0A5E4EFQ5|A0A5E4EFQ5	0.0
tr|A0A5E4GAW6|A0A5E4GAW6:tr|A0A5E4GBD6|A0A5E4GBD6:tr|F6K5V5|F6K5V5_9ROSA:tr|A0A5E4GAH5|A0A5E4GAH5	Alcohol dehydrogenase	26	1.72E+07	tr|A0A5E4GAW6|A0A5E4GAW6	0.0
tr|A0A4Y1RRK2|A0A4Y1RRK2:tr|A0A5E4GGE5|A0A5E4GGE5	Cobalamin-independent synthase family protein	15	1.65E+07	tr|A0A4Y1RRK2|A0A4Y1RRK2	0.0

tr|A0A5E4EZH8|A0A5E4EZH8	PREDICTED: heat shock	20	1.44E+07	tr|A0A5E4EZH8|A0A5E4EZH8	0.0
tr|A0A5E4EQW0|A0A5E4EQW0	ATP synthase subunit beta	28	1.44E+07	tr|A0A5E4EQW0|A0A5E4EQW0	0.0
tr|A0A5E4FHL8|A0A5E4FHL8	Tr-type G domain-containing protein	13	1.37E+07	tr|A0A5E4FHL8|A0A5E4FHL8	0.0
tr|A0A0P0CLD1|A0A0P0CLD1:tr|A0A4Y1RML5|A0A4Y1RML5:tr|A0A5E4EWS7|A0A5E4EWS7:tr|A0A4Y1RHW9|A0A4Y1RHW9:tr|A0A5E4EC29|A0A5E4EC29	Actin	21	1.35E+07	tr|A0A0P0CLD1|A0A0P0CLD1	0.0
tr|A0A5E4F2S4|A0A5E4F2S4	PREDICTED: leucine aminopeptidase	22	1.32E+07	tr|A0A5E4F2S4|A0A5E4F2S4	0.0
tr|A0A5E4F837|A0A5E4F837	PREDICTED: lysosomal	10	1.31E+07	tr|A0A5E4F837|A0A5E4F837	0.0
tr|A0A5E4G620|A0A5E4G620:tr|A0A4Y1R4J5|A0A4Y1R4J5	Late embryogenesis abundant protein family protein	35	1.21E+07	tr|A0A5E4G620|A0A5E4G620	0.0
tr|A0A5E4EGD4|A0A5E4EGD4:tr|A0A4Y1QTX0|A0A4Y1QTX0	Aspartate aminotransferase	22	1.21E+07	tr|A0A5E4EGD4|A0A5E4EGD4	0.0
tr|A0A5E4EVI7|A0A5E4EVI7	PREDICTED: heat shock	15	1.18E+07	tr|A0A5E4EVI7|A0A5E4EVI7	0.0
tr|A0A5H2XTX9|A0A5H2XTX9	ATP synthase subunit alpha	22	1.10E+07	tr|A0A5H2XTX9|A0A5H2XTX9	0.0
tr|A0A5E4EFT6|A0A5E4EFT6	PREDICTED: phosphoglucomutase	24	1.08E+07	tr|A0A5E4EFT6|A0A5E4EFT6	0.0
tr|A0A5H2XW79|A0A5H2XW79:tr|A0A5C2A5K3|A0A5C2A5K3:tr|A0A1W6CB98|A0A1W6CB98:tr|A0A2U8T536|A0A2U8T536:tr|A0A4Y1RFG0|A0A4Y1RFG0:tr|A0A5E4GQD1|A0A5E4GQD1	Ribulose bisphosphate carboxylase large chain	22	1.08E+07	tr|A0A5H2XW79|A0A5H2XW79	0.0
tr|A0A4Y1QSL3|A0A4Y1QSL3	Chaperone protein htpG family protein	15	1.07E+07	tr|A0A4Y1QSL3|A0A4Y1QSL3	0.0
tr|Q8H6U4|Q8H6U4:tr|A0A5E4EKI9|A0A5E4EKI9:tr|A0A4Y1S284|A0A4Y1S284	Heat shock protein 60	19	1.02E+07	tr|Q8H6U4|Q8H6U4	0.0
tr|A0A4Y1RFE1|A0A4Y1RFE1:tr|A0A5E4FY33|A0A5E4FY33:tr|A0A5E4FY40|A0A5E4FY40	Catalase	12	1.00E+07	tr|A0A4Y1RFE1|A0A4Y1RFE1	0.0
tr|A0A5E4G9K9|A0A5E4G9K9	UTP--glucose-1-phosphate uridylyltransferase	31	9.01E+06	tr|A0A5E4G9K9|A0A5E4G9K9	0.0
tr|A0A4Y1QMT2|A0A4Y1QMT2:tr|A0A5E4FBE8|A0A5E4FBE8	Sucrose synthase	6	8.86E+06	tr|A0A4Y1QMT2|A0A4Y1QMT2	0.0
tr|A0A5E4F297|A0A5E4F297:tr|A0A5E4F1H0|A0A5E4F1H0	Fructose-bisphosphate aldolase	25	8.60E+06	tr|A0A5E4F297|A0A5E4F297	0.0
tr|A0A5E4GEY2|A0A5E4GEY2	Oleosin	26	8.45E+06	tr|A0A5E4GEY2|A0A5E4GEY2	0.0
tr|A0A4Y1QYW9|A0A4Y1QYW9	Pyruvate kinase	26	8.36E+06	tr|A0A4Y1QYW9|A0A4Y1QYW9	0.0

tr|A0A5E4EBX5|A0A5E4EBX5	Urease	10	7.99E+06	tr|A0A5E4EBX5|A0A5E4EBX5	0.0
tr|A0A5E4FXZ0|A0A5E4FXZ0	Isocitrate lyase	21	7.34E+06	tr|A0A5E4FXZ0|A0A5E4FXZ0	0.0
tr|A0A5E4FH45|A0A5E4FH45	PREDICTED: putative	41	7.08E+06	tr|A0A5E4FH45|A0A5E4FH45	0.0
tr|A0A5E4FI64|A0A5E4FI64	PREDICTED: ATP-dependent	17	6.44E+06	tr|A0A5E4FI64|A0A5E4FI64	0.0
tr|A0A5E4GGR8|A0A5E4GGR8:tr|A0A5E4GFU5|A0A5E4GFU5	Xylose isomerase	15	6.43E+06	tr|A0A5E4GGR8|A0A5E4GGR8	0.0
tr|Q9SW89|Q9SW89	Abscisic acid response protein	26	6.28E+06	tr|Q9SW89|Q9SW89	0.0
tr|A0A5E4G542|A0A5E4G542:tr|A0A4Y1QNM7|A0A4Y1QNM7	Adenosylhomocysteinase	11	5.92E+06	tr|A0A5E4G542|A0A5E4G542	0.0
tr|A0A5E4F917|A0A5E4F917	PREDICTED: elongation	16	5.86E+06	tr|A0A5E4F917|A0A5E4F917	0.0
tr|A0A4Y1R1F0|A0A4Y1R1F0:tr|A0A5E4EBX0|A0A5E4EBX0:tr|A0A5E4EBU9|A0A5E4EBU9	Beta-hexosaminidase	9	5.71E+06	tr|A0A4Y1R1F0|A0A4Y1R1F0	0.0
tr|A0A5E4EK53|A0A5E4EK53	PREDICTED: aspartate--tRNA ligase	19	5.54E+06	tr|A0A5E4EK53|A0A5E4EK53	0.0
tr|A0A5E4FF40|A0A5E4FF40:tr|A0A5E4FRG1|A0A5E4FRG1	PREDICTED: lysM domain-containing GPI-anchored	17	5.43E+06	tr|A0A5E4FF40|A0A5E4FF40	0.0
tr|A0A5E4FKH8|A0A5E4FKH8:tr|A0A5E4EVI1|A0A5E4EVI1:tr|A0A4Y1R7Y3|A0A4Y1R7Y3:tr|A0A4Y1QR61|A0A4Y1QR61	PREDICTED: UDP-arabinopyranose mutase	17	5.42E+06	tr|A0A5E4FKH8|A0A5E4FKH8	0.0
tr|A0A5E4EQ45|A0A5E4EQ45	PREDICTED: betaine aldehyde dehydrogenase	16	5.15E+06	tr|A0A5E4EQ45|A0A5E4EQ45	0.0
tr|A0A5E4G5I2|A0A5E4G5I2:tr|A0A4Y1QNK9|A0A4Y1QNK9	Serine hydroxymethyltransferase	14	4.94E+06	tr|A0A5E4G5I2|A0A5E4G5I2	0.0
tr|A0A5E4E3T8|A0A5E4E3T8	PREDICTED: embryonic	15	4.89E+06	tr|A0A5E4E3T8|A0A5E4E3T8	0.0
tr|A0A5E4F6M0|A0A5E4F6M0	PREDICTED: oleosin	32	4.58E+06	tr|A0A5E4F6M0|A0A5E4F6M0	0.0
tr|A0A5E4E9Y8|A0A5E4E9Y8:tr|A0A5E4ECG2|A0A5E4ECG2	PREDICTED: phosphomannomutase/phosphoglucomutase	9	4.54E+06	tr|A0A5E4E9Y8|A0A5E4E9Y8	0.0
tr|A0A4Y1R3Q7|A0A4Y1R3Q7:tr|A0A5E4EY83|A0A5E4EY83	Alanine aminotransferase 2	14	4.31E+06	tr|A0A4Y1R3Q7|A0A4Y1R3Q7	0.0
tr|A0A5E4F541|A0A5E4F541:tr|A0A4Y1QSS4|A0A4Y1QSS4	Glucose-6-phosphate isomerase	14	4.29E+06	tr|A0A5E4F541|A0A5E4F541	0.0
tr|A0A5E4ECP0|A0A5E4ECP0	PREDICTED: LOW QUALITY PROTEIN	10	3.86E+06	tr|A0A5E4ECP0|A0A5E4ECP0	0.0
tr|A0A5E4FZ94|A0A5E4FZ94	PREDICTED: heat shock	11	3.74E+06	tr|A0A5E4FZ94|A0A5E4FZ94	0.0

tr|A0A5E4EKU2|A0A5E4EKU2	Glucose-6-phosphate isomerase	8	3.44E+06	tr|A0A5E4EKU2|A0A5E4EKU2	0.0
tr|A0A5E4F9T4|A0A5E4F9T4:tr|A0A4Y1R660|A0A4Y1R660	PREDICTED: C-1-tetrahydrofolate synthase	10	3.33E+06	tr|A0A5E4F9T4|A0A5E4F9T4	0.0
tr|A0A4Y1QSC2|A0A4Y1QSC2:tr|A0A5E4FCU6|A0A5E4FCU6:tr|A0A5E4FCR7|A0A5E4FCR7	TCP-1/cpn60 chaperonin family protein	9	3.26E+06	tr|A0A4Y1QSC2|A0A4Y1QSC2	0.0
tr|A0A5E4FYX0|A0A5E4FYX0	PREDICTED: phosphoenolpyruvate	8	2.53E+06	tr|A0A5E4FYX0|A0A5E4FYX0	0.0
tr|A0A5E4FPY3|A0A5E4FPY3	Pyruvate kinase	16	2.48E+06	tr|A0A5E4FPY3|A0A5E4FPY3	0.0
tr|A0A5E4EHV3|A0A5E4EHV3:tr|A0A5E4EC08|A0A5E4EC08	PREDICTED: probable nucleoredoxin	19	2.47E+06	tr|A0A5E4EHV3|A0A5E4EHV3	0.0
tr|A0A5E4F6U8|A0A5E4F6U8	Protein disulfide-isomerase	9	2.39E+06	tr|A0A5E4F6U8|A0A5E4F6U8	0.0
tr|A0A5E4F3G2|A0A5E4F3G2:tr|A0A4Y1RXW1|A0A4Y1RXW1	PREDICTED: heat shock	9	2.27E+06	tr|A0A5E4F3G2|A0A5E4F3G2	0.0
tr|A0A5E4EC35|A0A5E4EC35:tr|A0A4Y1R1Q4|A0A4Y1R1Q4	PREDICTED: triosephosphate isomerase	25	2.23E+06	tr|A0A5E4EC35|A0A5E4EC35	0.0
tr|A0A5E4FRK2|A0A5E4FRK2	Glutathione peroxidase	22	2.14E+06	tr|A0A5E4FRK2|A0A5E4FRK2	0.0
tr|A0A5E4FXJ2|A0A5E4FXJ2:tr|A0A4Y1R585|A0A4Y1R585	PREDICTED: T-complex	13	2.01E+06	tr|A0A5E4FXJ2|A0A5E4FXJ2	0.0
tr|A0A5E4FWP8|A0A5E4FWP8	PREDICTED: 14-3-3	26	1.91E+06	tr|A0A5E4FWP8|A0A5E4FWP8	0.0
tr|A0A5E4FU86|A0A5E4FU86:tr|A0A4Y1R9B2|A0A4Y1R9B2	ATP-dependent 6-phosphofructokinase	12	1.47E+06	tr|A0A5E4FU86|A0A5E4FU86	0.0
tr|A0A5E4EM19|A0A5E4EM19	Peptidyl-prolyl cis-trans isomerase	29	1.17E+06	tr|A0A5E4EM19|A0A5E4EM19	0.0

Table 8.S3. Identification of AEP2 band from SDS-PAGE by LC-MS/MS analysis of tryptic peptides produced by in-gel protein digestion.

Protein accession	Protein name	Coverage	Area	Protein accession for coverage and area	Relative abundance (%)
tr|A0A5E4EZP4|A0A5E4EZP4:tr|A0A4Y1QPK2|A0A4Y1QPK2	PREDICTED: vicilin	60	4.17E+10	tr|A0A5E4EZP4|A0A5E4EZP4	81.0
tr|A0A5E4FFS0|A0A5E4FFS0:sp|E3SH28|PRU01:sp|Q43607|PRU1	Prunin 1	73	6.96E+09	sp|Q43607|PRU1	13.5
tr|Q43608|Q43608:tr|E3SH29|E3SH29:tr|A0A5E4FK23|A0A5E4FK23:tr|A0A4Y1S2I9|A0A4Y1S2I9	Prunin 2	44	7.42E+08	tr|E3SH29|E3SH29	1.4

tr|A0A5E4GEN6|A0A5E4GEN6:sp|Q945K2|MDL2	(R)-mandelonitrile lyase 2	36	6.55E+08	sp|Q945K2|MDL2	1.3
tr|A0A4Y1QPI1|A0A4Y1QPI1:tr|A0A5E4F2T7|A0A5E4F2T7:tr|A0A5J6V1A4|A0A5J6V1A4	PREDICTED: vicilin	35	6.18E+08	tr|A0A4Y1QPI1|A0A4Y1QPI1	1.2
tr|A0A5E4ED32|A0A5E4ED32	Protein disulfide-isomerase	39	1.74E+08	tr|A0A5E4ED32|A0A5E4ED32	0.3
tr|A0A5E4EE27|A0A5E4EE27	PREDICTED: vicilin	33	1.60E+08	tr|A0A5E4EE27|A0A5E4EE27	0.3
tr|A0A5E4EAH6|A0A5E4EAH6	PREDICTED: basic 7S globulin	52	6.23E+07	tr|A0A5E4EAH6|A0A5E4EAH6	0.1
tr|A0A5E4G2Q4|A0A5E4G2Q4	PREDICTED: seed	27	5.97E+07	tr|A0A5E4G2Q4|A0A5E4G2Q4	0.1
tr|A0A5E4FV72|A0A5E4FV72	PREDICTED: vicilin	35	4.59E+07	tr|A0A5E4FV72|A0A5E4FV72	0.1
tr|A0A5E4FFP9|A0A5E4FFP9	Succinate-semialdehyde dehydrogenase	33	4.46E+07	tr|A0A5E4FFP9|A0A5E4FFP9	0.1
tr|A0A4Y1RVW6|A0A4Y1RVW6:tr|A0A5E4E244|A0A5E4E244	RmlC-like cupins superfamily protein	14	3.47E+07	tr|A0A4Y1RVW6|A0A4Y1RVW6	0.1
tr|A0A5E4FUY2|A0A5E4FUY2:tr|A0A5E4FLL1|A0A5E4FLL1:tr|A0A5H2XQ79|A0A5H2XQ79:tr|A0A5E4EET2|A0A5E4EET2	Elongation factor 1-alpha	25	3.14E+07	tr|A0A5E4FUY2|A0A5E4FUY2	0.1
tr|A0A5H2Y344|A0A5H2Y344:tr|A0A5E4ECQ1|A0A5E4ECQ1	NAD(P)-binding Rossmann-fold superfamily protein	39	2.51E+07	tr|A0A5H2Y344|A0A5H2Y344	0.0
tr|A0A4Y1QVV0|A0A4Y1QVV0:sp|P33629|TBA:tr|A0A5E4EJF0|A0A5E4EJF0:tr|A0A5E4EWJ5|A0A5E4EWJ5:tr|A0A5E4F747|A0A5E4F747:tr|A0A5E4F5H0|A0A5E4F5H0:tr|A0A4Y1S0R4|A0A4Y1S0R4	Tubulin alpha chain	11	2.38E+07	tr|A0A4Y1QVV0|A0A4Y1QVV0	0.0
sp|Q43804|OLEO1:tr|A0A5E4EAT1|A0A5E4EAT1	Oleosin 1	35	1.80E+07	sp|Q43804|OLEO1	0.0
tr|A0A5E4EYX0|A0A5E4EYX0:tr|A0A5E4EYT9|A0A5E4EYT9:tr|A0A516F3L2|A0A516F3L2	Cysteine rich antimicrobial protein	31	1.70E+07	tr|A0A5E4EYX0|A0A5E4EYX0	0.0
tr|A0A5E4F4B3|A0A5E4F4B3:tr|A0A5E4GGU0|A0A5E4GGU0:tr|A0A5E4GHG6|A0A5E4GHG6	PREDICTED: heat shock	28	1.30E+07	tr|A0A5E4F4B3|A0A5E4F4B3	0.0
tr|A0A5E4E8M3|A0A5E4E8M3	PREDICTED: 17	15	1.15E+07	tr|A0A5E4E8M3|A0A5E4E8M3	0.0
tr|A0A5E4E522|A0A5E4E522	PREDICTED: probable serine protease	12	1.10E+07	tr|A0A5E4E522|A0A5E4E522	0.0
tr|A0A5E4E5K6|A0A5E4E5K6:tr|A0A5E4E767|A0A5E4E767	PREDICTED: tripeptidyl-peptidase 2	11	1.06E+07	tr|A0A5E4E5K6|A0A5E4E5K6	0.0
tr|A0A5E4F917|A0A5E4F917	PREDICTED: elongation	22	8.43E+06	tr|A0A5E4F917|A0A5E4F917	0.0
tr|A0A4Y1R8N9|A0A4Y1R8N9	Alpha-mannosidase	10	8.35E+06	tr|A0A4Y1R8N9|A0A4Y1R8N9	0.0

tr|A0A5H2XW79|A0A5H2XW79:tr|A0A1W6CB98|A0A1W6CB98:tr|A0A2U8T597|A0A2U8T597:tr|A0A2U8T536|A0A2U8T536:tr|A0A5C2A5K3|A0A5C2A5K3:tr|A0A4Y1RFG0|A0A4Y1RFG0:tr|A0A5E4GQD1|A0A5E4GQD1	Ribulose bisphosphate carboxylase large chain	20	7.06E+06	tr|A0A5H2XW79|A0A5H2XW79	0.0
tr|H9ZGE2|H9ZGE2:tr|H9ZGD8|H9ZGD8:tr|A0A5H2XK07|A0A5H2XK07	Prunasin hydrolase	12	6.96E+06	tr|H9ZGE2|H9ZGE2	0.0
tr|A0A5E4ET55|A0A5E4ET55	Oleosin	30	6.83E+06	tr|A0A5E4ET55|A0A5E4ET55	0.0
tr|A0A4Y1R1F0|A0A4Y1R1F0:tr|A0A5E4EBX0|A0A5E4EBX0:tr|A0A5E4EBU9|A0A5E4EBU9	Beta-hexosaminidase	8	6.03E+06	tr|A0A4Y1R1F0|A0A4Y1R1F0	0.0
tr|A0A5E4EYP9|A0A5E4EYP9	PREDICTED: 11-beta-hydroxysteroid	19	5.98E+06	tr|A0A5E4EYP9|A0A5E4EYP9	0.0
tr|A0A5E4G542|A0A5E4G542:tr|A0A4Y1QNM7|A0A4Y1QNM7	Adenosylhomocysteinase	12	5.47E+06	tr|A0A5E4G542|A0A5E4G542	0.0
tr|A0A4Y1R8R9|A0A4Y1R8R9	Glycosyl hydrolase family 38 protein	9	5.02E+06	tr|A0A4Y1R8R9|A0A4Y1R8R9	0.0
tr|A0A4Y1RH78|A0A4Y1RH78:tr|A0A5E4F1W6|A0A5E4F1W6	Hydroxysteroid dehydrogenase 1	12	4.97E+06	tr|A0A4Y1RH78|A0A4Y1RH78	0.0
tr|A0A5E4G9K9|A0A5E4G9K9	UTP--glucose-1-phosphate uridylyltransferase	20	4.97E+06	tr|A0A5E4G9K9|A0A5E4G9K9	0.0
tr|A0A5E4FEI8|A0A5E4FEI8:tr|A0A4Y1R416|A0A4Y1R416:tr|A0A5E4E8P3|A0A5E4E8P3:tr|A0A5E4E3D0|A0A5E4E3D0:tr|A0A4Y1RV52|A0A4Y1RV52	Eukaryotic translation initiation factor 4A1	18	3.96E+06	tr|A0A5E4FEI8|A0A5E4FEI8	0.0
tr|A0A5E4G6M4|A0A5E4G6M4:tr|A0A5H2XK06|A0A5H2XK06	Late embryogenesis abundant protein family protein	15	3.65E+06	tr|A0A5E4G6M4|A0A5E4G6M4	0.0
tr|A0A4Y1QXP6|A0A4Y1QXP6	Eukaryotic aspartyl protease family protein	11	3.51E+06	tr|A0A4Y1QXP6|A0A4Y1QXP6	0.0
tr|A0A4Y1RTK3|A0A4Y1RTK3	Enolase	14	3.44E+06	tr|A0A4Y1RTK3|A0A4Y1RTK3	0.0
tr|A0A5E4GAW6|A0A5E4GAW6:tr|A0A5E4GBD6|A0A5E4GBD6:tr|F6K5V5|F6K5V5_9ROSA:tr|A0A5E4GAH5|A0A5E4GAH5	Alcohol dehydrogenase	13	3.16E+06	tr|A0A5E4GAW6|A0A5E4GAW6	0.0
tr|A0A5E4EGD4|A0A5E4EGD4:tr|A0A4Y1QTX0|A0A4Y1QTX0	Aspartate aminotransferase	13	2.72E+06	tr|A0A5E4EGD4|A0A5E4EGD4	0.0
tr|Q8H6U4|Q8H6U4:tr|A0A5E4EKI9|A0A5E4EKI9:tr|A0A4Y1S284|A0A4Y1S284	Heat shock protein 60	13	2.59E+06	tr|Q8H6U4|Q8H6U4	0.0
tr|A0A5E4ER27|A0A5E4ER27:tr|A0A4Y1RJK8|A0A4Y1RJK8	PREDICTED: aldehyde dehydrogenase	10	1.70E+06	tr|A0A5E4ER27|A0A5E4ER27	0.0
tr|A0A5E4G620|A0A5E4G620:tr|A0A4Y1R4J5|A0A4Y1R4J5	Late embryogenesis abundant protein family protein	25	1.31E+06	tr|A0A5E4G620|A0A5E4G620	0.0

tr|A0A5E4FTY1|A0A5E4FTY1	Annexin	16	1.23E+06	tr|A0A5E4FTY1|A0A5E4FTY1	0.0

Table 8.S4. Identification of AEP3 band from SDS-PAGE by LC-MS/MS analysis of tryptic peptides produced by in-gel protein digestion.

Protein accession	Protein name	Coverage	Area	Protein accession for coverage and area	Relative abundance (%)
tr|A0A5E4FFS0|A0A5E4FFS0:sp|E3SH28|PRU01:sp|Q43607|PRU1	Prunin 1	86	8.33E+10	sp|E3SH28|PRU01	93.5
tr|E3SH29|E3SH29:tr|Q43608|Q43608:tr|A0A4Y1S2I9|A0A4Y1S2I9:tr|A0A5E4FK23|A0A5E4FK23	Prunin 2	59	2.57E+09	tr|E3SH29|E3SH29	2.9
tr|A0A5E4EZP4|A0A5E4EZP4:tr|A0A4Y1QPK2|A0A4Y1QPK2	PREDICTED: vicilin	50	2.29E+09	tr|A0A5E4EZP4|A0A5E4EZP4	2.6
tr|A0A5E4GEN6|A0A5E4GEN6:sp|Q945K2|MDL2	(R)-mandelonitrile lyase 2	36	3.85E+08	sp|Q945K2|MDL2	0.4
tr|A0A4Y1QPI1|A0A4Y1QPI1:tr|A0A5E4F2T7|A0A5E4F2T7:tr|A0A5J6V1A4|A0A5J6V1A4	PREDICTED: vicilin	22	1.21E+08	tr|A0A4Y1QPI1|A0A4Y1QPI1	0.1
tr|A0A5E4F1W6|A0A5E4F1W6	PREDICTED: 11-beta-hydroxysteroid	46	5.97E+07	tr|A0A5E4F1W6|A0A5E4F1W6	0.1
tr|A0A5E4EE27|A0A5E4EE27	PREDICTED: vicilin	27	5.21E+07	tr|A0A5E4EE27|A0A5E4EE27	0.1
tr|A0A5E4EAH6|A0A5E4EAH6	PREDICTED: basic 7S globulin	53	5.19E+07	tr|A0A5E4EAH6|A0A5E4EAH6	0.1
tr|A0A5E4EYP9|A0A5E4EYP9	PREDICTED: 11-beta-hydroxysteroid	32	4.61E+07	tr|A0A5E4EYP9|A0A5E4EYP9	0.1
tr|A0A4Y1QTX0|A0A4Y1QTX0:tr|A0A5E4EGD4|A0A5E4EGD4	Aspartate aminotransferase	42	4.27E+07	tr|A0A4Y1QTX0|A0A4Y1QTX0	0.0
tr|A0A5E4FIM4|A0A5E4FIM4:tr|A0A4Y1RBA0|A0A4Y1RBA0	Phosphoglycerate kinase	38	2.99E+07	tr|A0A5E4FIM4|A0A5E4FIM4	0.0
tr|A0A5E4FV72|A0A5E4FV72	PREDICTED: vicilin	34	2.28E+07	tr|A0A5E4FV72|A0A5E4FV72	0.0
tr|A0A5E4GAW6|A0A5E4GAW6:tr|F6K5V5|F6K5V5_9ROSA:tr|A0A5E4GAH5|A0A5E4GAH5	Alcohol dehydrogenase	36	2.08E+07	tr|A0A5E4GAW6|A0A5E4GAW6	0.0
tr|A0A5E4F4B3|A0A5E4F4B3:tr|A0A5E4GHG6|A0A5E4GHG6:tr|A0A5E4GGU0|A0A5E4GGU0:tr|A0A5E4E9G9|A0A5E4E9G9	PREDICTED: heat shock	30	1.69E+07	tr|A0A5E4F4B3|A0A5E4F4B3	0.0
tr|A0A5E4FKH8|A0A5E4FKH8	PREDICTED: UDP-arabinopyranose mutase	33	1.47E+07	tr|A0A5E4FKH8|A0A5E4FKH8	0.0
tr|A0A516F3L2|A0A516F3L2:tr|A0A5E4EYT9|A0A5E4EYT9:tr|A0A5E4EYX0|A0A5E4EYX0	Cysteine rich antimicrobial protein	36	1.38E+07	tr|A0A516F3L2|A0A516F3L2	0.0

tr|A0A5H2Y344|A0A5H2Y344:tr|A0A5E4ECQ1|A0A5E4ECQ1	NAD(P)-binding Rossmann-fold superfamily protein	39	1.34E+07	tr|A0A5H2Y344|A0A5H2Y344	0.0
tr|A0A5E4G2Q4|A0A5E4G2Q4	PREDICTED: seed	19	1.23E+07	tr|A0A5E4G2Q4|A0A5E4G2Q4	0.0
tr|A0A4Y1RVW6|A0A4Y1RVW6:tr|A0A5E4E244|A0A5E4E244	RmlC-like cupins superfamily protein	17	1.18E+07	tr|A0A4Y1RVW6|A0A4Y1RVW6	0.0
tr|A0A5E4F297|A0A5E4F297	Fructose-bisphosphate aldolase	26	8.46E+06	tr|A0A5E4F297|A0A5E4F297	0.0
tr|A0A5E4ED32|A0A5E4ED32	Protein disulfide-isomerase	22	7.51E+06	tr|A0A5E4ED32|A0A5E4ED32	0.0
sp|Q43804|OLEO1	Oleosin 1	35	6.63E+06	sp|Q43804|OLEO1	0.0
tr|A0A5E4GHS7|A0A5E4GHS7	PREDICTED: SNF1-related kinase regulatory	14	5.29E+06	tr|A0A5E4GHS7|A0A5E4GHS7	0.0
tr|A0A5E4FFP9|A0A5E4FFP9	Succinate-semialdehyde dehydrogenase	15	4.78E+06	tr|A0A5E4FFP9|A0A5E4FFP9	0.0
tr|A0A5E4ERW0|A0A5E4ERW0:tr|A0A4Y1RCF0|A0A4Y1RCF0	PREDICTED: DJ-1	23	3.80E+06	tr|A0A5E4ERW0|A0A5E4ERW0	0.0
tr|A0A4Y1QQ02|A0A4Y1QQ02	Glyceraldehyde-3-phosphate dehydrogenase	18	2.10E+06	tr|A0A4Y1QQ02|A0A4Y1QQ02	0.0
tr|A0A5E4ES77|A0A5E4ES77	Formate dehydrogenase mitochondrial	11	1.88E+06	tr|A0A5E4ES77|A0A5E4ES77	0.0
tr|A0A4Y1RTK3|A0A4Y1RTK3	Enolase	16	1.70E+06	tr|A0A4Y1RTK3|A0A4Y1RTK3	0.0
tr|A0A4Y1R8R9|A0A4Y1R8R9	Glycosyl hydrolase family 38 protein	7	1.44E+06	tr|A0A4Y1R8R9|A0A4Y1R8R9	0.0
tr|A0A5E4G6M4|A0A5E4G6M4	PREDICTED: late embryogenesis abundant	17	1.35E+06	tr|A0A5E4G6M4|A0A5E4G6M4	0.0
tr|A0A4Y1R8N9|A0A4Y1R8N9	Alpha-mannosidase	8	1.17E+06	tr|A0A4Y1R8N9|A0A4Y1R8N9	0.0

Table 8.S5. Identification of AEP4 band from SDS-PAGE by LC-MS/MS analysis of tryptic peptides produced by in-gel protein digestion.

Protein accession	Protein name	Coverage	Area	Protein accession for coverage and area	Relative abundance (%)
tr|A0A5E4FFS0|A0A5E4FFS0:sp|Q43607|PRU1:sp|E3SH28|PRU01	Prunin 1	83	8.25E+10	sp|Q43607|PRU1	89.2

tr|E3SH29|E3SH29:tr|Q43608|Q43608:tr|A0A4Y1S2I9|A0A4Y1S2I9:tr|A0A5E4FK23|A0A5E4FK23	Prunin 2	66	6.56E+09	tr|E3SH29|E3SH29	7.1
tr|A0A5E4EZP4|A0A5E4EZP4:tr|A0A4Y1QPK2|A0A4Y1QPK2	PREDICTED: vicilin	46	1.35E+09	tr|A0A5E4EZP4|A0A5E4EZP4	1.5
tr|A0A4Y1QPI1|A0A4Y1QPI1	PREDICTED: vicilin	24	9.99E+08	tr|A0A4Y1QPI1|A0A4Y1QPI1	1.1
tr|A0A5E4GEN6|A0A5E4GEN6:sp|Q945K2|MDL2	(R)-mandelonitrile lyase 2	35	4.45E+08	sp|Q945K2|MDL2	0.5
tr|A0A5E4EYP9|A0A5E4EYP9	PREDICTED: 11-beta-hydroxysteroid	38	1.63E+08	tr|A0A5E4EYP9|A0A5E4EYP9	0.2
tr|A0A5E4EE27|A0A5E4EE27	PREDICTED: vicilin	32	1.12E+08	tr|A0A5E4EE27|A0A5E4EE27	0.1
tr|A0A4Y1QQ02|A0A4Y1QQ02:tr|A0A5E4F2W9|A0A5E4F2W9	Glyceraldehyde-3-phosphate dehydrogenase	30	4.99E+07	tr|A0A4Y1QQ02|A0A4Y1QQ02	0.1
tr|A0A5E4FV72|A0A5E4FV72	PREDICTED: vicilin	26	3.84E+07	tr|A0A5E4FV72|A0A5E4FV72	0.0
tr|A0A5H2Y344|A0A5H2Y344:tr|A0A5E4ECQ1|A0A5E4ECQ1	NAD(P)-binding Rossmann-fold superfamily protein	36	2.62E+07	tr|A0A5H2Y344|A0A5H2Y344	0.0
tr|A0A4Y1RH78|A0A4Y1RH78:tr|A0A5E4F1W6|A0A5E4F1W6	Hydroxysteroid dehydrogenase 1	22	2.38E+07	tr|A0A4Y1RH78|A0A4Y1RH78	0.0
tr|A0A5E4FKH8|A0A5E4FKH8:tr|A0A4Y1RIA9|A0A4Y1RIA9:tr|A0A5E4EVI1|A0A5E4EVI1:tr|A0A4Y1R7Y3|A0A4Y1R7Y3	PREDICTED: UDP-arabinopyranose mutase	22	2.04E+07	tr|A0A5E4FKH8|A0A5E4FKH8	0.0
tr|A0A5E4F4B3|A0A5E4F4B3	Uncharacterized protein	21	1.96E+07	tr|A0A5E4F4B3|A0A5E4F4B3	0.0
tr|A0A5E4EGD4|A0A5E4EGD4:tr|A0A4Y1QTX0|A0A4Y1QTX0	Aspartate aminotransferase	22	1.90E+07	tr|A0A5E4EGD4|A0A5E4EGD4	0.0
tr|A0A5E4G6M4|A0A5E4G6M4:tr|A0A5H2XK06|A0A5H2XK06	Late embryogenesis abundant protein family protein	28	1.89E+07	tr|A0A5E4G6M4|A0A5E4G6M4	0.0
tr|A0A5E4ET55|A0A5E4ET55	Oleosin	30	1.88E+07	tr|A0A5E4ET55|A0A5E4ET55	0.0
tr|A0A5E4ED32|A0A5E4ED32	Protein disulfide-isomerase	23	1.84E+07	tr|A0A5E4ED32|A0A5E4ED32	0.0
sp|Q43804|OLEO1:tr|A0A5E4EAT1|A0A5E4EAT1	Oleosin 1	35	1.46E+07	sp|Q43804|OLEO1	0.0
tr|A0A5E4G2Q4|A0A5E4G2Q4	PREDICTED: seed	19	1.44E+07	tr|A0A5E4G2Q4|A0A5E4G2Q4	0.0
tr|A0A5E4EYT9|A0A5E4EYT9:tr|A0A5E4EYX0|A0A5E4EYX0:tr|A0A516F3L2|A0A516F3L2	Cysteine rich antimicrobial protein	31	1.40E+07	tr|A0A5E4EYT9|A0A5E4EYT9	0.0
tr|A0A4Y1RY11|A0A4Y1RY11	Zinc-binding dehydrogenase family protein	8	1.32E+07	tr|A0A4Y1RY11|A0A4Y1RY11	0.0
tr|A0A5E4EK54|A0A5E4EK54	Malate dehydrogenase	21	1.04E+07	tr|A0A5E4EK54|A0A5E4EK54	0.0
tr|A0A5E4FAG3|A0A5E4FAG3	PREDICTED: actin	16	1.01E+07	tr|A0A5E4FAG3|A0A5E4FAG3	0.0

tr|A0A5E4F297|A0A5E4F297:tr|A0A5E4F1H0|A0A5E4F1H0:tr|A0A4Y1RTL3|A0A4Y1RTL3	Fructose-bisphosphate aldolase	25	9.79E+06	tr|A0A5E4F297|A0A5E4F297	0.0
tr|A0A5E4E8M3|A0A5E4E8M3:tr|A0A5E4E7H8|A0A5E4E7H8	PREDICTED: 17	13	8.98E+06	tr|A0A5E4E8M3|A0A5E4E8M3	0.0
tr|A0A5E4FFP9|A0A5E4FFP9	Succinate-semialdehyde dehydrogenase	12	8.39E+06	tr|A0A5E4FFP9|A0A5E4FFP9	0.0
tr|A0A5E4FIM4|A0A5E4FIM4:tr|A0A4Y1RBA0|A0A4Y1RBA0	Phosphoglycerate kinase	16	8.23E+06	tr|A0A5E4FIM4|A0A5E4FIM4	0.0
tr|A0A5E4E244|A0A5E4E244:tr|A0A4Y1RVW6|A0A4Y1RVW6	PREDICTED: vicilin	14	6.89E+06	tr|A0A5E4E244|A0A5E4E244	0.0
tr|A0A0P0CLD1|A0A0P0CLD1:tr|A0A4Y1RML5|A0A4Y1RML5:tr|A0A4Y1RHW9|A0A4Y1RHW9:tr|A0A5E4EWS7|A0A5E4EWS7:tr|A0A5E4FAG3|A0A5E4FAG3:tr|A0A5E4F5R8|A0A5E4F5R8:tr|A0A4Y1RKE9|A0A4Y1RKE9:tr|A0A5E4EC29|A0A5E4EC29	Actin	19	6.51E+06	tr|A0A0P0CLD1|A0A0P0CLD1	0.0
tr|A0A5E4GAW6|A0A5E4GAW6:tr|A0A5E4GBD6|A0A5E4GBD6:tr|F6K5V5|F6K5V5_9ROSA:tr|A0A5E4GAH5|A0A5E4GAH5	Alcohol dehydrogenase	12	6.01E+06	tr|A0A5E4GAW6|A0A5E4GAW6	0.0
tr|A0A5E4E824|A0A5E4E824	PREDICTED: serpin-ZX	17	5.85E+06	tr|A0A5E4E824|A0A5E4E824	0.0
tr|A0A4Y1R8N9|A0A4Y1R8N9	Alpha-mannosidase	11	5.19E+06	tr|A0A4Y1R8N9|A0A4Y1R8N9	0.0
tr|A0A5E4G0M9|A0A5E4G0M9	Cysteine synthase	14	3.66E+06	tr|A0A5E4G0M9|A0A5E4G0M9	0.0
tr|A0A5E4FTY1|A0A5E4FTY1	Annexin	18	1.72E+06	tr|A0A5E4FTY1|A0A5E4FTY1	0.0

Table 8.S6. Identification of AEP5 band from SDS-PAGE by LC-MS/MS analysis of tryptic peptides produced by in-gel protein digestion.

Protein accession	Protein name	Coverage	Area	Protein accession for coverage and area	Relative abundance (%)
sp|E3SH28|PRU01:tr|A0A5E4FFS0|A0A5E4FFS0:sp|Q43607|PRU1	Prunin 1	89	7.88E+10	sp|E3SH28|PRU01	86.4
tr|A0A5E4EE27|A0A5E4EE27	PREDICTED: vicilin	51	5.00E+09	tr|A0A5E4EE27|A0A5E4EE27	5.5
tr|A0A5E4EZP4|A0A5E4EZP4:tr|A0A4Y1QPK2|A0A4Y1QPK2	PREDICTED: vicilin	48	2.42E+09	tr|A0A5E4EZP4|A0A5E4EZP4	2.7
tr|E3SH29|E3SH29:tr|Q43608|Q43608:tr|A0A4Y1S2I9|A0A4Y1S2I9:tr|A0A5E4FK23|A0A5E4FK23	Prunin 2	68	2.03E+09	tr|E3SH29|E3SH29	2.2
tr|A0A4Y1QPI1|A0A4Y1QPI1:tr|A0A5E4F2T7|A0A5E4F2T7:tr|A0A5J6V1A4|A0A5J6V1A4	PREDICTED: vicilin	28	1.07E+09	tr|A0A4Y1QPI1|A0A4Y1QPI1	1.2
tr|A0A5H2Y344|A0A5H2Y344:tr|A0A5E4ECQ1|A0A5E4ECQ1:tr|A0A5E4ECQ0|A0A5E4ECQ0	NAD(P)-binding Rossmann-fold superfamily protein	65	7.70E+08	tr|A0A5H2Y344|A0A5H2Y344	0.8

tr|A0A5E4GEN6|A0A5E4GEN6:sp|Q945K2|MDL2	(R)-mandelonitrile lyase 2	34	2.82E+08	sp|Q945K2|MDL2	0.3
tr|A0A5E4FV72|A0A5E4FV72	PREDICTED: vicilin	30	1.87E+08	tr|A0A5E4FV72|A0A5E4FV72	0.2
tr|A0A5E4GE42|A0A5E4GE42:tr|A0A4Y1QME7|A0A4Y1QME7	Uncharacterized protein	41	1.59E+08	tr|A0A5E4GE42|A0A5E4GE42	0.2
tr|A0A5E4FTY1|A0A5E4FTY1	Annexin	55	9.97E+07	tr|A0A5E4FTY1|A0A5E4FTY1	0.1
tr|A0A5E4EYP9|A0A5E4EYP9	PREDICTED: 11-beta-hydroxysteroid	23	5.23E+07	tr|A0A5E4EYP9|A0A5E4EYP9	0.1
tr|A0A5E4EYX0|A0A5E4EYX0:tr|A0A5E4EYT9|A0A5E4EYT9:tr|A0A516F3L2|A0A516F3L2	Cysteine rich antimicrobial protein	47	4.49E+07	tr|A0A5E4EYX0|A0A5E4EYX0	0.0
tr|A0A4Y1QQ02|A0A4Y1QQ02:tr|A0A5E4F2W9|A0A5E4F2W9	Glyceraldehyde-3-phosphate dehydrogenase	26	3.42E+07	tr|A0A4Y1QQ02|A0A4Y1QQ02	0.0
tr|A0A5E4ET55|A0A5E4ET55	Oleosin	26	2.29E+07	tr|A0A5E4ET55|A0A5E4ET55	0.0
sp|Q43804|OLEO1:tr|A0A5E4EAT1|A0A5E4EAT1	Oleosin 1	37	2.28E+07	sp|Q43804|OLEO1	0.0
tr|A0A5E4FVK4|A0A5E4FVK4	PREDICTED: isoflavone reductase	19	2.17E+07	tr|A0A5E4FVK4|A0A5E4FVK4	0.0
tr|A0A5E4FUY2|A0A5E4FUY2:tr|A0A5E4EET2|A0A5E4EET2	Elongation factor 1-alpha	14	1.98E+07	tr|A0A5E4FUY2|A0A5E4FUY2	0.0
tr|A0A5E4FN81|A0A5E4FN81:tr|A0A5E4FN74|A0A5E4FN74	Annexin	39	1.85E+07	tr|A0A5E4FN81|A0A5E4FN81	0.0
tr|F6K5V5|F6K5V5_9ROSA:tr|A0A5E4GAW6|A0A5E4GAW6:tr|A0A5E4GBD6|A0A5E4GBD6:tr|A0A5E4GAH5|A0A5E4GAH5:tr|A0A5E4G1H1|A0A5E4G1H1	Alcohol dehydrogenase	18	1.65E+07	tr|F6K5V5|F6K5V5_9ROSA	0.0
tr|A0A5E4E7H8|A0A5E4E7H8:tr|A0A5E4E8M3|A0A5E4E8M3	PREDICTED: 17	14	1.46E+07	tr|A0A5E4E7H8|A0A5E4E7H8	0.0
tr|A0A5E4EAH6|A0A5E4EAH6	PREDICTED: basic 7S globulin	33	1.41E+07	tr|A0A5E4EAH6|A0A5E4EAH6	0.0
tr|A0A5E4F4B3|A0A5E4F4B3:tr|A0A5E4GGU0|A0A5E4GGU0:tr|A0A5E4GHG6|A0A5E4GHG6	PREDICTED: heat shock	26	1.37E+07	tr|A0A5E4F4B3|A0A5E4F4B3	0.0
tr|A0A5E4EX25|A0A5E4EX25:tr|A0A4Y1R7Q6|A0A4Y1R7Q6	General regulatory factor 3	24	1.28E+07	tr|A0A5E4EX25|A0A5E4EX25	0.0
tr|A0A5E4FWP8|A0A5E4FWP8	PREDICTED: 14-3-3	34	1.24E+07	tr|A0A5E4FWP8|A0A5E4FWP8	0.0
tr|A0A5E4F1W6|A0A5E4F1W6	PREDICTED: 11-beta-hydroxysteroid	37	1.19E+07	tr|A0A5E4F1W6|A0A5E4F1W6	0.0
tr|A0A5E4F729|A0A5E4F729	Lactoylglutathione lyase	23	1.13E+07	tr|A0A5E4F729|A0A5E4F729	0.0
tr|A0A5E4ED32|A0A5E4ED32	Protein disulfide-isomerase	21	1.10E+07	tr|A0A5E4ED32|A0A5E4ED32	0.0
tr|A0A5E4G6M4|A0A5E4G6M4:tr|A0A5H2XK06|A0A5H2XK06	Late embryogenesis abundant protein family protein	22	1.09E+07	tr|A0A5E4G6M4|A0A5E4G6M4	0.0

tr|A0A5E4G2Q4|A0A5E4G2Q4	PREDICTED: seed	29	8.85E+06	tr|A0A5E4G2Q4|A0A5E4G2Q4	0.0
tr|A0A5E4E244|A0A5E4E244	PREDICTED: vicilin	13	8.36E+06	tr|A0A5E4E244|A0A5E4E244	0.0
tr|A0A4Y1R8N9|A0A4Y1R8N9	Alpha-mannosidase	9	7.46E+06	tr|A0A4Y1R8N9|A0A4Y1R8N9	0.0
tr|A0A5E4FAF3|A0A5E4FAF3:tr|A0A4Y1R9Z0|A0A4Y1R9Z0	PKS_ER domain-containing protein	21	7.16E+06	tr|A0A5E4FAF3|A0A5E4FAF3	0.0
tr|A0A5E4FPJ9|A0A5E4FPJ9	PREDICTED: DUF1264	18	6.86E+06	tr|A0A5E4FPJ9|A0A5E4FPJ9	0.0
tr|A0A5E4FX06|A0A5E4FX06:tr|A0A4Y1RFZ0|A0A4Y1RFZ0:tr|A0A5E4FWS0|A0A5E4FWS0	Beta vacuolar processing enzyme	15	6.44E+06	tr|A0A5E4FX06|A0A5E4FX06	0.0
tr|A0A5E4G0M9|A0A5E4G0M9	Cysteine synthase	17	4.27E+06	tr|A0A5E4G0M9|A0A5E4G0M9	0.0
tr|A0A0P0CLD1|A0A0P0CLD1:tr|A0A4Y1RML5|A0A4Y1RML5:tr|A0A4Y1RKE9|A0A4Y1RKE9:tr|A0A5E4F5R8|A0A5E4F5R8:tr|A0A5E4FAG3|A0A5E4FAG3	Actin	19	4.02E+06	tr|A0A0P0CLD1|A0A0P0CLD1	0.0
tr|A0A5E4F297|A0A5E4F297:tr|A0A4Y1RTL3|A0A4Y1RTL3:tr|A0A5E4F1H0|A0A5E4F1H0	Fructose-bisphosphate aldolase	18	3.46E+06	tr|A0A5E4F297|A0A5E4F297	0.0
tr|Q9SW89|Q9SW89	Abscisic acid response protein	32	2.78E+06	tr|Q9SW89|Q9SW89	0.0
tr|A0A5E4E5K6|A0A5E4E5K6:tr|A0A5E4E767|A0A5E4E767	PREDICTED: tripeptidyl-peptidase 2	5	2.51E+06	tr|A0A5E4E5K6|A0A5E4E5K6	0.0
tr|A0A5E4FIM4|A0A5E4FIM4:tr|A0A4Y1RBA0|A0A4Y1RBA0	Phosphoglycerate kinase	12	1.39E+06	tr|A0A5E4FIM4|A0A5E4FIM4	0.0
tr|A0A5E4FFP9|A0A5E4FFP9	Succinate-semialdehyde dehydrogenase	10	1.32E+06	tr|A0A5E4FFP9|A0A5E4FFP9	0.0

Table 8.S7. Identification of AEP6 band from SDS-PAGE by LC-MS/MS analysis of tryptic peptides produced by in-gel protein digestion.

Protein accession	Protein name	Coverage	Area	Protein accession for coverage and area	Relative abundance (%)
tr|A0A5E4FFS0|A0A5E4FFS0:sp|E3SH28|PRU01:sp|Q43607|PRU1	Prunin 1	82	3.87E+10	sp|E3SH28|PRU01	72.8
tr|A0A5E4EE27|A0A5E4EE27	PREDICTED: vicilin	56	5.19E+09	tr|A0A5E4EE27|A0A5E4EE27	9.8
tr|Q43608|Q43608:tr|E3SH29|E3SH29:tr|A0A5E4FK23|A0A5E4FK23:tr|A0A4Y1S2I9|A0A4Y1S2I9	Prunin 2	56	2.61E+09	tr|E3SH29|E3SH29	4.9
tr|A0A5E4FV72|A0A5E4FV72	PREDICTED: vicilin	41	1.93E+09	tr|A0A5E4FV72|A0A5E4FV72	3.6
tr|A0A5E4EZP4|A0A5E4EZP4:tr|A0A4Y1QPK2|A0A4Y1QPK2	PREDICTED: vicilin	48	1.07E+09	tr|A0A5E4EZP4|A0A5E4EZP4	2.0
tr|A0A4Y1QPI1|A0A4Y1QPI1:tr|A0A5E4F2T7|A0A5E4F2T7:tr|A0A5J6V1A4|A0A5J6V1A4	PREDICTED: vicilin	35	8.93E+08	tr|A0A4Y1QPI1|A0A4Y1QPI1	1.7

tr|A0A5E4ECQ1|A0A5E4ECQ1:tr|A0A5H2Y344|A0A5H2Y344	NAD(P)-binding Rossmann-fold superfamily protein	65	5.34E+08	tr|A0A5E4ECQ1|A0A5E4ECQ1	1.0
tr|A0A5E4GEN6|A0A5E4GEN6:sp|Q945K2|MDL2	(R)-mandelonitrile lyase 2	34	4.44E+08	sp|Q945K2|MDL2	0.8
tr|A0A4Y1QME7|A0A4Y1QME7:tr|A0A5E4GE42|A0A5E4GE42	Uncharacterized protein	36	2.12E+08	tr|A0A4Y1QME7|A0A4Y1QME7	0.4
sp|Q43804|OLEO1:tr|A0A5E4EAT1|A0A5E4EAT1	Oleosin 1	39	1.67E+08	sp|Q43804|OLEO1	0.3
tr|A0A5E4EYX0|A0A5E4EYX0:tr|A0A5E4EYT9|A0A5E4EYT9:tr|A0A516F3L2|A0A516F3L2	Cysteine rich antimicrobial protein	50	1.26E+08	tr|A0A5E4EYX0|A0A5E4EYX0	0.2
tr|A0A5E4GEY2|A0A5E4GEY2	Oleosin	40	1.19E+08	tr|A0A5E4GEY2|A0A5E4GEY2	0.2
tr|A0A5E4EKE0|A0A5E4EKE0	PREDICTED: peroxygenase	44	8.68E+07	tr|A0A5E4EKE0|A0A5E4EKE0	0.2
tr|A0A5E4ERY7|A0A5E4ERY7	Superoxide dismutase	30	8.65E+07	tr|A0A5E4ERY7|A0A5E4ERY7	0.2
tr|A0A5E4ET55|A0A5E4ET55	Oleosin	36	8.52E+07	tr|A0A5E4ET55|A0A5E4ET55	0.2
tr|A0A5E4EAH6|A0A5E4EAH6	PREDICTED: basic 7S globulin	45	6.05E+07	tr|A0A5E4EAH6|A0A5E4EAH6	0.1
tr|A0A4Y1QQ02|A0A4Y1QQ02:tr|A0A5E4F2W9|A0A5E4F2W9	Glyceraldehyde-3-phosphate dehydrogenase	31	5.68E+07	tr|A0A4Y1QQ02|A0A4Y1QQ02	0.1
tr|A0A5E4E8M3|A0A5E4E8M3	PREDICTED: 17	18	5.40E+07	tr|A0A5E4E8M3|A0A5E4E8M3	0.1
tr|A0A5H2XRV0|A0A5H2XRV0:tr|A0A5E4EFV9|A0A5E4EFV9:tr|A0A4Y1RS80|A0A4Y1RS80	Glycoprotein membrane GPI-anchored protein	11	4.80E+07	tr|A0A5H2XRV0|A0A5H2XRV0	0.1
tr|A0A5E4EP81|A0A5E4EP81	PREDICTED: superoxide dismutase	22	4.56E+07	tr|A0A5E4EP81|A0A5E4EP81	0.1
tr|A0A5E4FWQ4|A0A5E4FWQ4	Uncharacterized protein	25	4.51E+07	tr|A0A5E4FWQ4|A0A5E4FWQ4	0.1
tr|F6K5V5|F6K5V5_9ROSA:tr|A0A5E4GAW6|A0A5E4GAW6:tr|A0A5E4GBD6|A0A5E4GBD6:tr|A0A5E4GAH5|A0A5E4GAH5	Alcohol dehydrogenase	39	3.83E+07	tr|F6K5V5|F6K5V5_9ROSA	0.1
tr|A0A5E4EYP9|A0A5E4EYP9	PREDICTED: 11-beta-hydroxysteroid	19	3.57E+07	tr|A0A5E4EYP9|A0A5E4EYP9	0.1
tr|A0A5E4FPJ9|A0A5E4FPJ9	PREDICTED: DUF1264	43	3.41E+07	tr|A0A5E4FPJ9|A0A5E4FPJ9	0.1
tr|A0A5E4GKU6|A0A5E4GKU6	PREDICTED: TenA family	14	3.32E+07	tr|A0A5E4GKU6|A0A5E4GKU6	0.1
tr|A0A4Y1R1Q4|A0A4Y1R1Q4:tr|A0A5E4EC35|A0A5E4EC35	Triosephosphate isomerase	26	2.94E+07	tr|A0A4Y1R1Q4|A0A4Y1R1Q4	0.1
tr|A0A5E4G2Q4|A0A5E4G2Q4	PREDICTED: seed	40	2.81E+07	tr|A0A5E4G2Q4|A0A5E4G2Q4	0.1
tr|A0A4Y1R1F6|A0A4Y1R1F6	Dehydroascorbate reductase 2	50	2.57E+07	tr|A0A4Y1R1F6|A0A4Y1R1F6	0.0
tr|A0A5E4FNY9|A0A5E4FNY9	PREDICTED: glutathione	17	2.51E+07	tr|A0A5E4FNY9|A0A5E4FNY9	0.0

tr|A0A5E4F1W6|A0A5E4F1W6:tr|A0A4Y1RH78|A0A4Y1RH78	PREDICTED: 11-beta-hydroxysteroid	35	2.36E+07	tr|A0A5E4F1W6|A0A5E4F1W6	0.0
tr|A0A5E4F5S2|A0A5E4F5S2:tr|A0A5E4F5X4|A0A5E4F5X4	Amine oxidase	8	2.25E+07	tr|A0A5E4F5S2|A0A5E4F5S2	0.0
tr|A0A4Y1QMF8|A0A4Y1QMF8	1-cysteine peroxiredoxin 1	51	2.21E+07	tr|A0A4Y1QMF8|A0A4Y1QMF8	0.0
tr|A0A5E4G6M4|A0A5E4G6M4:tr|A0A5H2XK06|A0A5H2XK06	Late embryogenesis abundant protein family protein	24	2.12E+07	tr|A0A5E4G6M4|A0A5E4G6M4	0.0
tr|A0A5E4FUY2|A0A5E4FUY2:tr|A0A5E4EET2|A0A5E4EET2	Elongation factor 1-alpha	15	1.99E+07	tr|A0A5E4FUY2|A0A5E4FUY2	0.0
tr|A0A5E4F4B3|A0A5E4F4B3:tr|A0A5E4E9G9|A0A5E4E9G9	PREDICTED: heat shock	28	1.70E+07	tr|A0A5E4F4B3|A0A5E4F4B3	0.0
tr|A0A4Y1R9V5|A0A4Y1R9V5:tr|A0A5E4F729|A0A5E4F729	Lactoylglutathione lyase	30	1.69E+07	tr|A0A4Y1R9V5|A0A4Y1R9V5	0.0
tr|A0A5E4FTY1|A0A5E4FTY1	Annexin	44	1.67E+07	tr|A0A5E4FTY1|A0A5E4FTY1	0.0
tr|A0A5E4EX25|A0A5E4EX25:tr|A0A4Y1R7Q6|A0A4Y1R7Q6	PREDICTED: 14-3-3	32	1.67E+07	tr|A0A5E4EX25|A0A5E4EX25	0.0
tr|A0A4Y1R8N9|A0A4Y1R8N9	Alpha-mannosidase	11	1.62E+07	tr|A0A4Y1R8N9|A0A4Y1R8N9	0.0
tr|A0A5E4FFP9|A0A5E4FFP9	Succinate-semialdehyde dehydrogenase	14	1.39E+07	tr|A0A5E4FFP9|A0A5E4FFP9	0.0
tr|A0A5E4EH09|A0A5E4EH09:tr|A0A5E4E4E5|A0A5E4E4E5:tr|A0A5E4EI60|A0A5E4EI60:tr|A0A5E4E5X3|A0A5E4E5X3:tr|A0A4Y1QUG8|A0A4Y1QUG8	GTP-binding nuclear protein	32	1.36E+07	tr|A0A5E4EH09|A0A5E4EH09	0.0
tr|A0A5E4F6U9|A0A5E4F6U9	PREDICTED: thaumatin	19	1.28E+07	tr|A0A5E4F6U9|A0A5E4F6U9	0.0
tr|A0A5E4E244|A0A5E4E244	PREDICTED: vicilin	13	1.28E+07	tr|A0A5E4E244|A0A5E4E244	0.0
tr|A0A5E4EGD4|A0A5E4EGD4:tr|A0A4Y1QTX0|A0A4Y1QTX0	Aspartate aminotransferase	23	1.18E+07	tr|A0A5E4EGD4|A0A5E4EGD4	0.0
tr|A0A5E4E824|A0A5E4E824	PREDICTED: serpin-ZX	16	1.09E+07	tr|A0A5E4E824|A0A5E4E824	0.0
tr|H9ZGE2|H9ZGE2:tr|H9ZGD8|H9ZGD8:tr|A0A5H2XK07|A0A5H2XK07:tr|A0A4Y1RMD2|A0A4Y1RMD2:tr|A0A5E4GM82|A0A5E4GM82	Prunasin hydrolase	15	1.07E+07	tr|H9ZGE2|H9ZGE2	0.0
tr|A0A5E4EZ94|A0A5E4EZ94	PREDICTED: 14-3-3	25	9.78E+06	tr|A0A5E4EZ94|A0A5E4EZ94	0.0
tr|A0A4Y1RBA0|A0A4Y1RBA0:tr|A0A5E4FIM4|A0A5E4FIM4	Phosphoglycerate kinase	32	9.37E+06	tr|A0A4Y1RBA0|A0A4Y1RBA0	0.0
tr|A0A5E4FN81|A0A5E4FN81:tr|A0A5E4FN74|A0A5E4FN74	Annexin	32	9.13E+06	tr|A0A5E4FN81|A0A5E4FN81	0.0
tr|A0A5E4FQ12|A0A5E4FQ12	PREDICTED: stem-specific	25	8.65E+06	tr|A0A5E4FQ12|A0A5E4FQ12	0.0
tr|A0A5E4FX06|A0A5E4FX06:tr|A0A5E4FWS0|A0A5E4FWS0:tr|A0A4Y1RFZ0|A0A4Y1RFZ0	Beta vacuolar processing enzyme	15	8.43E+06	tr|A0A5E4FX06|A0A5E4FX06	0.0
tr|A0A5E4E5K6|A0A5E4E5K6:tr|A0A5E4E767|A0A5E4E767	PREDICTED: tripeptidyl-peptidase 2	12	7.32E+06	tr|A0A5E4E5K6|A0A5E4E5K6	0.0

tr|A0A4Y1RTL3|A0A4Y1RTL3:tr|A0A5E4F297|A0A5E4F297	Fructose-bisphosphate aldolase	13	6.80E+06	tr|A0A4Y1RTL3|A0A4Y1RTL3	0.0
tr|A0A5E4FKH8|A0A5E4FKH8:tr|A0A4Y1RIA9|A0A4Y1RIA9:tr|A0A5E4EVI1|A0A5E4EVI1:tr|A0A4Y1R7Y3|A0A4Y1R7Y3	PREDICTED: UDP-arabinopyranose mutase	15	6.80E+06	tr|A0A5E4FKH8|A0A5E4FKH8	0.0
tr|A0A5E4G9M9|A0A5E4G9M9:tr|A0A5H2XV04|A0A5H2XV04	Cysteine proteinase inhibitor	54	6.20E+06	tr|A0A5E4G9M9|A0A5E4G9M9	0.0
tr|A0A5E4F6M0|A0A5E4F6M0	PREDICTED: oleosin	30	4.77E+06	tr|A0A5E4F6M0|A0A5E4F6M0	0.0
tr|A0A5E4F250|A0A5E4F250	Proteasome subunit alpha type	22	4.62E+06	tr|A0A5E4F250|A0A5E4F250	0.0
tr|A0A4Y1RTK3|A0A4Y1RTK3	Enolase	14	4.57E+06	tr|A0A4Y1RTK3|A0A4Y1RTK3	0.0
tr|A0A5E4E3T8|A0A5E4E3T8	PREDICTED: embryonic	13	3.92E+06	tr|A0A5E4E3T8|A0A5E4E3T8	0.0
tr|A0A5E4FH45|A0A5E4FH45	PREDICTED: putative	41	3.87E+06	tr|A0A5E4FH45|A0A5E4FH45	0.0
tr|A0A5E4E522|A0A5E4E522	PREDICTED: probable serine protease	11	3.27E+06	tr|A0A5E4E522|A0A5E4E522	0.0
tr|A0A5E4EUV0|A0A5E4EUV0	PREDICTED: hydroxyacylglutathione hydrolase	20	3.10E+06	tr|A0A5E4EUV0|A0A5E4EUV0	0.0
tr|A0A5E4FBS8|A0A5E4FBS8	PREDICTED: NDR1/HIN1	28	2.80E+06	tr|A0A5E4FBS8|A0A5E4FBS8	0.0

Table 8.S8. Identification of AEP7 band from SDS-PAGE by LC-MS/MS analysis of tryptic peptides produced by in-gel protein digestion.

Protein accession	Protein name	Coverage	Area	Protein accession for coverage and area	Relative abundance (%)
sp|E3SH28|PRU01:tr|A0A5E4FFS0|A0A5E4FFS0:sp|Q43607|PRU1	Prunin 1	90	4.66E+11	sp|E3SH28|PRU01	93.8
tr|E3SH29|E3SH29:tr|Q43608|Q43608:tr|A0A5E4FK23|A0A5E4FK23:tr|A0A4Y1S2I9|A0A4Y1S2I9	Prunin 2	75	2.71E+10	tr|E3SH29|E3SH29	5.5
tr|A0A5E4EE27|A0A5E4EE27	PREDICTED: vicilin	48	1.82E+09	tr|A0A5E4EE27|A0A5E4EE27	0.4
tr|A0A5E4FV72|A0A5E4FV72	PREDICTED: vicilin	41	5.47E+08	tr|A0A5E4FV72|A0A5E4FV72	0.1
tr|A0A5E4EZP4|A0A5E4EZP4	PREDICTED: vicilin	41	4.06E+08	tr|A0A5E4EZP4|A0A5E4EZP4	0.1
tr|A0A5E4E8M3|A0A5E4E8M3	PREDICTED: 17	18	1.85E+08	tr|A0A5E4E8M3|A0A5E4E8M3	0.0
tr|A0A5E4GEN6|A0A5E4GEN6:sp|Q945K2|MDL2	(R)-mandelonitrile lyase 2	23	1.31E+08	sp|Q945K2|MDL2	0.0
tr|A0A5H2Y344|A0A5H2Y344:tr|A0A5E4ECQ1|A0A5E4ECQ1	NAD(P)-binding Rossmann-fold	37	7.69E+07	tr|A0A5H2Y344|A0A5H2Y344	0.0

	superfamily protein				
tr|A0A5E4E244|A0A5E4E244	PREDICTED: vicilin	18	6.61E+07	tr|A0A5E4E244| A0A5E4E244	0.0
tr|A0A5E4EYX0|A0A5E4EYX0:tr|A0A5E4EYT9|A0A5E4EYT9:tr|A0A516F3L2| A0A516F3L2	Cysteine rich antimicrobial protein	45	5.91E+07	tr|A0A5E4EYX0| A0A5E4EYX0	0.0
tr|A0A5E4GEY2|A0A5E4GEY2	Oleosin	26	5.72E+07	tr|A0A5E4GEY2| A0A5E4GEY2	0.0
tr|A0A5E4EAT1|A0A5E4EAT1	Oleosin	34	3.57E+07	tr|A0A5E4EAT1| A0A5E4EAT1	0.0
tr|A0A5E4EKE0|A0A5E4EKE0	PREDICTED: peroxygenase	25	2.45E+07	tr|A0A5E4EKE0| A0A5E4EKE0	0.0
tr|A0A5E4ET55|A0A5E4ET55	Oleosin	29	1.85E+07	tr|A0A5E4ET55| A0A5E4ET55	0.0
tr|A0A4Y1QPI1|A0A4Y1QPI1:tr|A0A5E4F2T7|A0A5E4F2T7:tr|A0A5J6V1A4|A0A5J6V1A4	PREDICTED: vicilin	9	1.42E+07	tr|A0A4Y1QPI1| A0A4Y1QPI1	0.0
tr|A0A5H2XPA3|A0A5H2XPA3:tr|A0A5E4FYT3|A0A5E4FYT3	HSP20-like chaperones superfamily protein	28	9.46E+06	tr|A0A5H2XPA3| A0A5H2XPA3	0.0
tr|A0A5E4FG68|A0A5E4FG68	PREDICTED: 22	18	6.77E+06	tr|A0A5E4FG68| A0A5E4FG68	0.0
tr|Q9SW89|Q9SW89	Abscisic acid response protein	31	6.29E+06	tr|Q9SW89|Q9SW89	0.0
tr|A0A5E4EPA1|A0A5E4EPA1	Uncharacterized protein	29	1.57E+06	tr|A0A5E4EPA1| A0A5E4EPA1	0.0

Table 8.S9. Identification of AEP8 band from SDS-PAGE by LC-MS/MS analysis of tryptic peptides produced by in-gel protein digestion.

Protein accession	Protein name	Coverage	Area	Protein accession for coverage and area	Relative abundance (%)
tr|E3SH29|E3SH29:tr|A0A5E4FK23|A0A5E4FK23:tr|A0A4Y1S2I9|A0A4Y1S2I9:tr|Q43608|Q43608	Prunin 2	72	2.02E+11	tr|E3SH29|E3SH29	78.6
sp|E3SH28|PRU01:tr|A0A5E4FFS0|A0A5E4FFS0:sp|Q43607|PRU1	Prunin 1	87	5.20E+10	sp|E3SH28|PRU01	20.3
tr|A0A5E4EZP4|A0A5E4EZP4	PREDICTED: vicilin	43	6.59E+08	tr|A0A5E4EZP4| A0A5E4EZP4	0.3
tr|A0A5E4EE27|A0A5E4EE27	PREDICTED: vicilin	46	3.84E+08	tr|A0A5E4EE27| A0A5E4EE27	0.1
tr|A0A5E4FV72|A0A5E4FV72	PREDICTED: vicilin	35	3.25E+08	tr|A0A5E4FV72| A0A5E4FV72	0.1
tr|A0A5E4EKE0|A0A5E4EKE0	PREDICTED: peroxygenase	32	2.40E+08	tr|A0A5E4EKE0| A0A5E4EKE0	0.1
tr|A0A5H2Y344|A0A5H2Y344:tr|A0A5E4ECQ1|A0A5E4ECQ1	NAD(P)-binding Rossmann-fold superfamily protein	46	1.80E+08	tr|A0A5H2Y344| A0A5H2Y344	0.1
tr|A0A4Y1RUU1|A0A4Y1RUU1:tr|A0A5E4EJV1|A0A5E4EJV1	Adenine nucleotide alpha hydrolases-like	46	1.49E+08	tr|A0A4Y1RUU1| A0A4Y1RUU1	0.1

	superfamily protein				
tr|A0A5E4GEY2|A0A5E4GEY2	Oleosin	37	1.48E+08	tr|A0A5E4GEY2|A0A5E4GEY2	0.1
tr|A0A5E4E7H8|A0A5E4E7H8:tr|A0A5E4E8M3|A0A5E4E8M3	PREDICTED: 17	17	1.19E+08	tr|A0A5E4E7H8|A0A5E4E7H8	0.0
tr|A0A4Y1RK38|A0A4Y1RK38:tr|A0A4Y1RK37|A0A4Y1RK37	HSP20-like chaperones superfamily protein	24	9.45E+07	tr|A0A4Y1RK38|A0A4Y1RK38	0.0
tr|A0A5E4E244|A0A5E4E244	PREDICTED: vicilin	19	8.70E+07	tr|A0A5E4E244|A0A5E4E244	0.0
tr|A0A5E4EYX0|A0A5E4EYX0:tr|A0A5E4EYT9|A0A5E4EYT9:tr|A0A516F3L2|A0A516F3L2	Cysteine rich antimicrobial protein	38	6.18E+07	tr|A0A5E4EYX0|A0A5E4EYX0	0.0
tr|A0A4Y1QPI1|A0A4Y1QPI1:tr|A0A5E4F2T7|A0A5E4F2T7:tr|A0A5J6V1A4|A0A5J6V1A4	PREDICTED: vicilin	19	5.42E+07	tr|A0A4Y1QPI1|A0A4Y1QPI1	0.0
sp|Q945K2|MDL2:tr|A0A5E4GEN6|A0A5E4GEN6	(R)-mandelonitrile lyase 2	25	4.95E+07	sp|Q945K2|MDL2	0.0
tr|A0A5E4EAT1|A0A5E4EAT1	Oleosin	35	3.29E+07	tr|A0A5E4EAT1|A0A5E4EAT1	0.0
tr|A0A5E4FA77|A0A5E4FA77	PREDICTED: 22	44	3.25E+07	tr|A0A5E4FA77|A0A5E4FA77	0.0
tr|Q9SW89|Q9SW89:tr|A0A5E4E399|A0A5E4E399	Abscisic acid response protein	33	2.87E+07	tr|Q9SW89|Q9SW89	0.0
tr|A0A5E4EES0|A0A5E4EES0	PREDICTED: YbhB/YbcL family Raf kinase inhibitor	26	2.01E+07	tr|A0A5E4EES0|A0A5E4EES0	0.0
tr|A0A5E4F6M0|A0A5E4F6M0	PREDICTED: oleosin	32	1.59E+07	tr|A0A5E4F6M0|A0A5E4F6M0	0.0
tr|A0A5E4ET55|A0A5E4ET55	Oleosin	29	1.47E+07	tr|A0A5E4ET55|A0A5E4ET55	0.0
tr|A0A5E4FRK2|A0A5E4FRK2	Glutathione peroxidase	27	1.38E+07	tr|A0A5E4FRK2|A0A5E4FRK2	0.0
tr|A0A5E4GJC3|A0A5E4GJC3:tr|A0A4Y1RGW0|A0A4Y1RGW0	HSP20-like chaperones superfamily protein	22	1.02E+07	tr|A0A5E4GJC3|A0A5E4GJC3	0.0
tr|A0A5E4E6S4|A0A5E4E6S4	PREDICTED: small	28	9.01E+06	tr|A0A5E4E6S4|A0A5E4E6S4	0.0
tr|A0A5E4FPJ9|A0A5E4FPJ9	PREDICTED: DUF1264	25	8.35E+06	tr|A0A5E4FPJ9|A0A5E4FPJ9	0.0
tr|A0A5E4G620|A0A5E4G620:tr|A0A4Y1R4J5|A0A4Y1R4J5	Late embryogenesis abundant protein family protein	47	6.12E+06	tr|A0A5E4G620|A0A5E4G620	0.0
tr|A0A5H2XK06|A0A5H2XK06:tr|A0A5E4G6M4|A0A5E4G6M4	Late embryogenesis abundant protein family protein	14	5.64E+06	tr|A0A5H2XK06|A0A5H2XK06	0.0
tr|A0A5E4EAH6|A0A5E4EAH6	PREDICTED: basic 7S globulin	18	3.53E+06	tr|A0A5E4EAH6|A0A5E4EAH6	0.0

tr|A0A5E4G2Q4|A0A5E4G2Q4	PREDICTED: seed	11	2.11E+06	tr|A0A5E4G2Q4|A0A5E4G2Q4	0.0

Table 8.S10. Identification of AEP9 band from SDS-PAGE by LC-MS/MS analysis of tryptic peptides produced by in-gel protein digestion.

Protein accession	Protein name	Coverage	Area	Protein accession for coverage and area	Relative abundance (%)
sp|E3SH28|PRU01:tr|A0A5E4FFS0|A0A5E4FFS0:sp|Q43607|PRU1	Prunin 1	84	3.15E+10	sp|E3SH28|PRU01	36.4
tr|A0A5E4EZP4|A0A5E4EZP4	PREDICTED: vicilin	66	1.14E+10	tr|A0A5E4EZP4|A0A5E4EZP4	13.2
tr|A0A5E4EYX0|A0A5E4EYX0:tr|A0A5E4EYT9|A0A5E4EYT9:tr|A0A516F3L2|A0A516F3L2	Cysteine rich antimicrobial protein	74	1.09E+10	tr|A0A5E4EYX0|A0A5E4EYX0	12.6
tr|E3SH29|E3SH29:tr|Q43608|Q43608:tr|A0A4Y1S2I9|A0A4Y1S2I9:tr|A0A5E4FK23|A0A5E4FK23	Prunin 2	74	1.06E+10	tr|E3SH29|E3SH29	12.3
tr|A0A5E4ET55|A0A5E4ET55	Oleosin	47	6.69E+09	tr|A0A5E4ET55|A0A5E4ET55	7.7
sp|Q43804|OLEO1:tr|A0A5E4EAT1|A0A5E4EAT1	Oleosin 1	42	6.07E+09	sp|Q43804|OLEO1	7.0
tr|A0A5E4FV72|A0A5E4FV72	PREDICTED: vicilin	39	1.17E+09	tr|A0A5E4FV72|A0A5E4FV72	1.3
tr|A0A5E4F2T7|A0A5E4F2T7:tr|A0A4Y1QPI1|A0A4Y1QPI1	PREDICTED: vicilin	41	1.03E+09	tr|A0A4Y1QPI1|A0A4Y1QPI1	1.2
tr|A0A4Y1RK38|A0A4Y1RK38	HSP20-like chaperones superfamily protein	17	8.18E+08	tr|A0A4Y1RK38|A0A4Y1RK38	0.9
tr|A0A4Y1RUU1|A0A4Y1RUU1:tr|A0A5E4EJV1|A0A5E4EJV1	Adenine nucleotide alpha hydrolases-like superfamily protein	61	7.11E+08	tr|A0A4Y1RUU1|A0A4Y1RUU1	0.8
tr|A0A5E4EE27|A0A5E4EE27	PREDICTED: vicilin	46	5.92E+08	tr|A0A5E4EE27|A0A5E4EE27	0.7
tr|A0A5E4GEY2|A0A5E4GEY2	Oleosin	40	5.21E+08	tr|A0A5E4GEY2|A0A5E4GEY2	0.6
tr|A0A5E4E7H8|A0A5E4E7H8:tr|A0A5E4E8M3|A0A5E4E8M3	PREDICTED: 17	30	5.07E+08	tr|A0A5E4E7H8|A0A5E4E7H8	0.6
tr|A0A5E4FEW8|A0A5E4FEW8	PREDICTED: DUF1264 domain-containing	25	4.59E+08	tr|A0A5E4FEW8|A0A5E4FEW8	0.5
tr|A0A5E4GEN6|A0A5E4GEN6:sp|Q945K2|MDL2	(R)-mandelonitrile lyase 2	43	4.12E+08	sp|Q945K2|MDL2	0.5
tr|Q9SW89|Q9SW89:tr|A0A5E4E399|A0A5E4E399	Abscisic acid response protein	56	2.79E+08	tr|Q9SW89|Q9SW89	0.3
tr|A0A4Y1RRI6|A0A4Y1RRI6	Non-specific lipid-transfer protein	10	2.13E+08	tr|A0A4Y1RRI6|A0A4Y1RRI6	0.2

tr|A0A5E4E244|A0A5E4E244	PREDICTED: vicilin	18	1.81E+08	tr|A0A5E4E244|A0A5E4E244	0.2
tr|A0A5E4FH45|A0A5E4FH45	PREDICTED: putative	67	1.80E+08	tr|A0A5E4FH45|A0A5E4FH45	0.2
tr|A0A4Y1R4J5|A0A4Y1R4J5:tr|A0A5E4G620|A0A5E4G620	Late embryogenesis abundant protein family protein	48	1.76E+08	tr|A0A4Y1R4J5|A0A4Y1R4J5	0.2
tr|A0A5H2Y344|A0A5H2Y344:tr|A0A5E4ECQ1|A0A5E4ECQ1	NAD(P)-binding Rossmann-fold superfamily protein	34	1.39E+08	tr|A0A5H2Y344|A0A5H2Y344	0.2
tr|A0A5E4F6M0|A0A5E4F6M0	PREDICTED: oleosin	32	1.26E+08	tr|A0A5E4F6M0|A0A5E4F6M0	0.1
tr|A0A4Y1QQ02|A0A4Y1QQ02:tr|A0A5E4F2W9|A0A5E4F2W9	Glyceraldehyde-3-phosphate dehydrogenase	37	1.15E+08	tr|A0A4Y1QQ02|A0A4Y1QQ02	0.1
tr|A0A5E4EKE0|A0A5E4EKE0	PREDICTED: peroxygenase	36	1.11E+08	tr|A0A5E4EKE0|A0A5E4EKE0	0.1
tr|A0A5E4EFS2|A0A5E4EFS2:tr|A0A4Y1RPW0|A0A4Y1RPW0	PREDICTED: em	52	9.53E+07	tr|A0A5E4EFS2|A0A5E4EFS2	0.1
tr|B5LXD0|B5LXD0	Small heat shock protein	79	8.28E+07	tr|B5LXD0|B5LXD0	0.1
tr|A0A5E4EAH6|A0A5E4EAH6	PREDICTED: basic 7S globulin	52	6.88E+07	tr|A0A5E4EAH6|A0A5E4EAH6	0.1
tr|A0A5E4E3T8|A0A5E4E3T8	PREDICTED: embryonic	24	6.86E+07	tr|A0A5E4E3T8|A0A5E4E3T8	0.1
tr|A0A5E4EZB7|A0A5E4EZB7:tr|A0A5E4FGS1|A0A5E4FGS1:tr|A0A4Y1RTB4|A0A4Y1RTB4:tr|A0A4Y1R2S3|A0A4Y1R2S3:tr|A0A5E4FP47|A0A5E4FP47:tr|A0A4Y1R0L8|A0A4Y1R0L8:tr|A0A4Y1RCH1|A0A4Y1RCH1:tr|A0A5E4F0U3|A0A5E4F0U3:tr|A0A5E4FKE9|A0A5E4FKE9	Ubiquitin supergroup	30	6.44E+07	tr|A0A5E4EZB7|A0A5E4EZB7	0.1
tr|A0A5E4G7L6|A0A5E4G7L6	PREDICTED: glycine-rich	58	6.19E+07	tr|A0A5E4G7L6|A0A5E4G7L6	0.1
tr|A0A5E4F472|A0A5E4F472:tr|A0A4Y1RDA8|A0A4Y1RDA8	Endoribonuclease L-PSP family protein	47	6.11E+07	tr|A0A5E4F472|A0A5E4F472	0.1
tr|A0A5E4F1C0|A0A5E4F1C0:tr|A0A5E4F0H8|A0A5E4F0H8:tr|A0A4Y1R9Q1|A0A4Y1R9Q1	Aspartic proteinase A1	21	5.82E+07	tr|A0A5E4F1C0|A0A5E4F1C0	0.1
tr|A0A5E4G6M4|A0A5E4G6M4:tr|A0A5H2XK06|A0A5H2XK06	Late embryogenesis abundant protein family protein	32	5.18E+07	tr|A0A5E4G6M4|A0A5E4G6M4	0.1
tr|A0A5E4GM30|A0A5E4GM30	PREDICTED: dehydrin	25	5.15E+07	tr|A0A5E4GM30|A0A5E4GM30	0.1
tr|A0A4Y1RWA1|A0A4Y1RWA1:tr|A0A5E4FA77|A0A5E4FA77	HSP20-like chaperones superfamily protein	32	4.70E+07	tr|A0A4Y1RWA1|A0A4Y1RWA1	0.1
tr|A0A5E4GAW6|A0A5E4GAW6:tr|F6K5V5|F6K5V5_9ROSA:tr|A0A5E4GAH5|A0A5E4GAH5	Alcohol dehydrogenase	43	4.29E+07	tr|A0A5E4GAW6|A0A5E4GAW6	0.0
tr|A0A5E4EWS7|A0A5E4EWS7:tr|A0A4Y1RHW9|A0A4Y1RHW9:tr|A0A0P0CL	Actin 7	26	3.59E+07	tr|A0A5E4EWS7|A0A5E4EWS7	0.0

D1|A0A0P0CLD1:tr|A0A4Y1RML5|A0A4Y1RML5:tr|A0A5E4FAG3|A0A5E4FAG3:tr|A0A5E4EC29|A0A5E4EC29:tr|A0A5E4F5R8|A0A5E4F5R8:tr|A0A4Y1RKE9|A0A4Y1RKE9					
tr|A0A5E4FIK2|A0A5E4FIK2	PREDICTED: nucleoside	36	3.58E+07	tr|A0A5E4FIK2|A0A5E4FIK2	0.0
tr|A0A5E4FKW1|A0A5E4FKW1	PREDICTED: universal stress	31	3.53E+07	tr|A0A5E4FKW1|A0A5E4FKW1	0.0
tr|A0A5E4FD94|A0A5E4FD94	PREDICTED: gibberellin-regulated	32	3.51E+07	tr|A0A5E4FD94|A0A5E4FD94	0.0
tr|A0A5E4GJC3|A0A5E4GJC3	PREDICTED: small	30	3.39E+07	tr|A0A5E4GJC3|A0A5E4GJC3	0.0
tr|A0A5E4F4B3|A0A5E4F4B3	Uncharacterized protein	27	3.18E+07	tr|A0A5E4F4B3|A0A5E4F4B3	0.0
tr|A0A5E4G2Q4|A0A5E4G2Q4	PREDICTED: seed	28	3.14E+07	tr|A0A5E4G2Q4|A0A5E4G2Q4	0.0
tr|A0A5E4F729|A0A5E4F729	Lactoylglutathione lyase	29	3.01E+07	tr|A0A5E4F729|A0A5E4F729	0.0
tr|A0A5E4F9Q4|A0A5E4F9Q4:tr|A0A5E4FC44|A0A5E4FC44	PREDICTED: major allergen	59	3.01E+07	tr|A0A5E4F9Q4|A0A5E4F9Q4	0.0
tr|A0A5E4FA48|A0A5E4FA48:tr|B6CQR8|B6CQR8_9ROSA:tr|B6CQR7|B6CQR7_9ROSA	PREDICTED: major allergen	40	2.95E+07	tr|A0A5E4FA48|A0A5E4FA48	0.0
tr|A0A5E4FTY1|A0A5E4FTY1	Annexin	42	2.85E+07	tr|A0A5E4FTY1|A0A5E4FTY1	0.0
tr|A0A5E4FPJ9|A0A5E4FPJ9	PREDICTED: DUF1264	25	2.69E+07	tr|A0A5E4FPJ9|A0A5E4FPJ9	0.0
tr|A0A5E4FRK2|A0A5E4FRK2	Glutathione peroxidase	26	2.69E+07	tr|A0A5E4FRK2|A0A5E4FRK2	0.0
tr|A0A4Y1QMR4|A0A4Y1QMR4	MLP-like protein 423	25	2.64E+07	tr|A0A4Y1QMR4|A0A4Y1QMR4	0.0
tr|A0A5E4ENE7|A0A5E4ENE7	PREDICTED: ribonuclease	41	2.59E+07	tr|A0A5E4ENE7|A0A5E4ENE7	0.0
tr|A0A5E4FSH3|A0A5E4FSH3	PREDICTED: glutaredoxin	45	2.32E+07	tr|A0A5E4FSH3|A0A5E4FSH3	0.0
tr|A0A5E4EM19|A0A5E4EM19	Peptidyl-prolyl cis-trans isomerase	47	2.23E+07	tr|A0A5E4EM19|A0A5E4EM19	0.0
tr|A0A5E4ERY7|A0A5E4ERY7:tr|A0A4Y1R213|A0A4Y1R213	Superoxide dismutase	24	2.13E+07	tr|A0A5E4ERY7|A0A5E4ERY7	0.0
tr|A0A5E4F8U2|A0A5E4F8U2	PREDICTED: late embryogenesis abundant	17	2.08E+07	tr|A0A5E4F8U2|A0A5E4F8U2	0.0
tr|A0A5E4F1W6|A0A5E4F1W6	PREDICTED: 11-beta-hydroxysteroid	33	1.99E+07	tr|A0A5E4F1W6|A0A5E4F1W6	0.0
tr|A0A4Y1QMF8|A0A4Y1QMF8	1-cysteine peroxiredoxin 1	35	1.80E+07	tr|A0A4Y1QMF8|A0A4Y1QMF8	0.0
tr|A0A5E4EAI1|A0A5E4EAI1	PREDICTED: 17	25	1.74E+07	tr|A0A5E4EAI1|A0A5E4EAI1	0.0
tr|A0A5E4FUY2|A0A5E4FUY2:tr|A0A5E4EET2|A0A5E4EET2	Elongation factor 1-alpha	14	1.63E+07	tr|A0A5E4FUY2|A0A5E4FUY2	0.0
tr|A0A5E4FYS3|A0A5E4FYS3	PREDICTED: embryo-specific	23	1.62E+07	tr|A0A5E4FYS3|A0A5E4FYS3	0.0

tr|A0A5E4FT33|A0A5E4FT33:tr|A0A5E4FTB3|A0A5E4FTB3	PREDICTED: cysteine	25	1.53E+07	tr|A0A5E4FT33|A0A5E4FT33	0.0
tr|H9ZGE3|H9ZGE3:tr|H9ZGD9|H9ZGD9	Prunasin hydrolase	17	1.39E+07	tr|H9ZGE3|H9ZGE3	0.0
tr|A0A4Y1RBA0|A0A4Y1RBA0:tr|A0A5E4FIM4|A0A5E4FIM4	Phosphoglycerate kinase	21	1.32E+07	tr|A0A4Y1RBA0|A0A4Y1RBA0	0.0
tr|A0A4Y1RMB4|A0A4Y1RMB4	Dessication-induced 1VOC superfamily protein	44	1.25E+07	tr|A0A4Y1RMB4|A0A4Y1RMB4	0.0
tr|A0A4Y1QTX0|A0A4Y1QTX0:tr|A0A5E4EGD4|A0A5E4EGD4	Aspartate aminotransferase	18	1.14E+07	tr|A0A4Y1QTX0|A0A4Y1QTX0	0.0
tr|A0A5H2XQR0|A0A5H2XQR0:tr|A0A5E4EQP7|A0A5E4EQP7:tr|A0A5E4G0B8|A0A5E4G0B8:tr|A0A5E4EEK3|A0A5E4EEK3:tr|A0A4Y1RRL4|A0A4Y1RRL4:tr|A0A5E4EPU0|A0A5E4EPU0:tr|A0A4Y1R874|A0A4Y1R874:tr|A0A4Y1QQE3|A0A4Y1QQE3	ADP-ribosylation factor A1F	27	9.07E+06	tr|A0A5H2XQR0|A0A5H2XQR0	0.0
tr|A0A5E4EFA6|A0A5E4EFA6:tr|A0A4Y1S383|A0A4Y1S383	PREDICTED: universal stress	36	8.68E+06	tr|A0A5E4EFA6|A0A5E4EFA6	0.0
tr|A0A5E4G829|A0A5E4G829:tr|A0A4Y1RJT1|A0A4Y1RJT1	Cystathionine beta-synthase family protein	20	8.65E+06	tr|A0A5E4G829|A0A5E4G829	0.0
tr|A0A4Y1RTK3|A0A4Y1RTK3	Enolase	17	8.52E+06	tr|A0A4Y1RTK3|A0A4Y1RTK3	0.0
tr|A0A5E4FFP9|A0A5E4FFP9	Succinate-semialdehyde dehydrogenase	12	7.99E+06	tr|A0A5E4FFP9|A0A5E4FFP9	0.0
tr|A0A5E4E754|A0A5E4E754	PREDICTED: transmembrane	29	7.90E+06	tr|A0A5E4E754|A0A5E4E754	0.0
tr|B6CQU6|B6CQU6_9ROSA:tr|A0A5E4EFD5|A0A5E4EFD5	Non-specific lipid-transfer protein	34	7.84E+06	tr|B6CQU6|B6CQU6_9ROSA	0.0
tr|A0A5E4F4S5|A0A5E4F4S5	PREDICTED: aspartic	10	6.98E+06	tr|A0A5E4F4S5|A0A5E4F4S5	0.0
tr|A0A4Y1RM49|A0A4Y1RM49	PREDICTED: 60S ribosomal	26	6.30E+06	tr|A0A4Y1RM49|A0A4Y1RM49	0.0
tr|A0A5E4EDN1|A0A5E4EDN1	PREDICTED: zinc ribbon domain-containing	29	4.22E+06	tr|A0A5E4EDN1|A0A5E4EDN1	0.0
tr|A0A4Y1QW77|A0A4Y1QW77	Plantacyanin	25	4.12E+06	tr|A0A4Y1QW77|A0A4Y1QW77	0.0
tr|A0A5E4EYP9|A0A5E4EYP9	PREDICTED: 11-beta-hydroxysteroid	18	3.72E+06	tr|A0A5E4EYP9|A0A5E4EYP9	0.0
tr|A0A5E4ER94|A0A5E4ER94:tr|A0A5H2XJX7|A0A5H2XJX7	Ribosomal protein S5 domain 2-like superfamily protein	41	3.18E+06	tr|A0A5E4ER94|A0A5E4ER94	0.0
tr|A0A5H2XPA3|A0A5H2XPA3:tr|A0A5E4FYT3|A0A5E4FYT3	HSP20-like chaperones superfamily protein	26	3.12E+06	tr|A0A5H2XPA3|A0A5H2XPA3	0.0

tr|A0A5E4GKZ1|A0A5E4GKZ1	PREDICTED: cytochrome	23	2.82E+06	tr|A0A5E4GKZ1|A0A5E4GKZ1	0.0
tr|A0A5E4EW53|A0A5E4EW53	PREDICTED: universal stress	27	2.64E+06	tr|A0A5E4EW53|A0A5E4EW53	0.0
tr|A0A5E4F0I8|A0A5E4F0I8:tr|A0A5H2Y1L1|A0A5H2Y1L1	Seed maturation protein	26	2.36E+06	tr|A0A5E4F0I8|A0A5E4F0I8	0.0
tr|A0A5E4E5K6|A0A5E4E5K6:tr|A0A5E4E767|A0A5E4E767	PREDICTED: tripeptidyl-peptidase 2	6	1.97E+06	tr|A0A5E4E5K6|A0A5E4E5K6	0.0
tr|A0A5E4F244|A0A5E4F244	PREDICTED: seed maturation PM41	29	1.81E+06	tr|A0A5E4F244|A0A5E4F244	0.0
tr|A0A5E4F503|A0A5E4F503:tr|A0A4Y1QLI6|A0A4Y1QLI6	PREDICTED: glutaredoxin-C4	34	1.71E+06	tr|A0A5E4F503|A0A5E4F503	0.0
tr|A0A5E4E6F1|A0A5E4E6F1	Peroxiredoxin	40	7.17E+05	tr|A0A5E4E6F1|A0A5E4E6F1	0.0

Table 8.S11. Identification of EAEP1 band from SDS-PAGE by LC-MS/MS analysis of tryptic peptides produced by in-gel protein digestion.

Protein accession	Protein name	Coverage	Area	Protein accession for coverage and area	Relative abundance (%)
tr|A0A5E4FFS0|A0A5E4FFS0:sp|E3SH28|PRU01:sp|Q43607|PRU1	Prunin 1	67	1.06E+10	sp|E3SH28|PRU01	63.8
tr|Q43608|Q43608:tr|A0A5E4FK23|A0A5E4FK23:tr|E3SH29|E3SH29:tr|A0A4Y1S2I9|A0A4Y1S2I9	Prunin 2	41	1.83E+09	tr|E3SH29|E3SH29	11.1
tr|A0A5E4EZP4|A0A5E4EZP4	PREDICTED: vicilin	43	1.18E+09	tr|A0A5E4EZP4|A0A5E4EZP4	7.1
tr|A0A4Y1QPI1|A0A4Y1QPI1:tr|A0A5E4F2T7|A0A5E4F2T7:tr|A0A5J6V1A4|A0A5J6V1A4	PREDICTED: vicilin	32	9.98E+08	tr|A0A4Y1QPI1|A0A4Y1QPI1	6.0
tr|A0A4Y1QME7|A0A4Y1QME7:tr|A0A5E4GE42|A0A5E4GE42	Uncharacterized protein	53	6.68E+08	tr|A0A4Y1QME7|A0A4Y1QME7	4.0
tr|A0A5E4FV72|A0A5E4FV72	PREDICTED: vicilin	35	3.40E+08	tr|A0A5E4FV72|A0A5E4FV72	2.1
tr|A0A5E4EE27|A0A5E4EE27	PREDICTED: vicilin	33	3.40E+08	tr|A0A5E4EE27|A0A5E4EE27	2.0
tr|A0A5E4GEN6|A0A5E4GEN6:sp|Q945K2|MDL2	(R)-mandelonitrile lyase 2	31	2.13E+08	sp|Q945K2|MDL2	1.3
tr|A0A5E4EKE0|A0A5E4EKE0	PREDICTED: peroxygenase	42	1.04E+08	tr|A0A5E4EKE0|A0A5E4EKE0	0.6
tr|A0A5H2XM09|A0A5H2XM09:tr|A0A5E4FGF6|A0A5E4FGF6:tr|A0A4Y1RB79|A0A4Y1RB79	Glyceraldehyde-3-phosphate dehydrogenase	15	6.95E+07	tr|A0A5H2XM09|A0A5H2XM09	0.4
tr|A0A5E4E244|A0A5E4E244	PREDICTED: vicilin	7	6.14E+07	tr|A0A5E4E244|A0A5E4E244	0.4
tr|A0A5E4EYX0|A0A5E4EYX0:tr|A0A5E4EYT9|A0A5E4EYT9:tr|A0A516F3L2|A0A516F3L2	Cysteine rich antimicrobial protein	37	4.25E+07	tr|A0A5E4EYX0|A0A5E4EYX0	0.3
tr|A0A4Y1RH78|A0A4Y1RH78:tr|A0A5E4F1W6|A0A5E4F1W6	Hydroxysteroid dehydrogenase 1	18	2.76E+07	tr|A0A4Y1RH78|A0A4Y1RH78	0.2

tr|A0A5E4FWQ4|A0A5E4FWQ4	Uncharacterized protein	21	2.18E+07	tr|A0A5E4FWQ4| A0A5E4FWQ4	0.1
tr|A0A5E4EYP9|A0A5E4EYP9:tr|A0A4Y 1QVS3|A0A4Y1QVS3	Hydroxysteroid dehydrogenase 5	17	2.00E+07	tr|A0A5E4EYP9| A0A5E4EYP9	0.1
tr|A0A5E4ERY7|A0A5E4ERY7	Superoxide dismutase	23	1.75E+07	tr|A0A5E4ERY7| A0A5E4ERY7	0.1
tr|A0A5E4ECQ1|A0A5E4ECQ1:tr|A0A5 H2Y344|A0A5H2Y344	NAD(P)-binding Rossmann-fold superfamily protein	23	1.47E+07	tr|A0A5E4ECQ1| A0A5E4ECQ1	0.1
tr|A0A5E4E824|A0A5E4E824	PREDICTED: serpin-ZX	25	1.39E+07	tr|A0A5E4E824| A0A5E4E824	0.1
tr|A0A5E4EAH6|A0A5E4EAH6	PREDICTED: basic 7S globulin	24	1.04E+07	tr|A0A5E4EAH6| A0A5E4EAH6	0.1
tr|A0A5E4E543|A0A5E4E543	Malic enzyme	9	5.19E+06	tr|A0A5E4E543| A0A5E4E543	0.0
tr|A0A5E4F6U9|A0A5E4F6U9	PREDICTED: thaumatin	23	4.17E+06	tr|A0A5E4F6U9| A0A5E4F6U9	0.0
tr|A0A4Y1RIB8|A0A4Y1RIB8:tr|A0A5E 4F1F2|A0A5E4F1F2	Malate synthase	5	3.94E+06	tr|A0A4Y1RIB8| A0A4Y1RIB8	0.0
sp|Q43804|OLEO1:tr|A0A5E4EAT1|A0A 5E4EAT1	Oleosin 1	34	3.32E+06	sp|Q43804|OLEO 1	0.0
tr|A0A5E4FH41|A0A5E4FH41:tr|A0A5E 4FH45|A0A5E4FH45	PREDICTED: putative	33	8.10E+05	tr|A0A5E4FH41| A0A5E4FH41	0.0

Table 8.S12. Identification of EAEP2 band from SDS-PAGE by LC-MS/MS analysis of tryptic peptides produced by in-gel protein digestion.

Protein accession	Protein name	Coverage	Area	Protein accession for coverage and area	Relative abundance (%)
tr|A0A5E4FFS0|A0A5E4FFS0:sp|E3SH28 |PRU01:sp|Q43607|PRU1	Prunin 1	59	2.49E+11	sp|E3SH28|PRU0 1	98.1
tr|E3SH29|E3SH29:tr|A0A5E4FK23|A0A 5E4FK23:tr|Q43608|Q43608:tr|A0A4Y1S 2I9|A0A4Y1S2I9	Prunin 2	48	3.42E+09	tr|E3SH29|E3SH2 9	1.3
tr|A0A5E4EZP4|A0A5E4EZP4	PREDICTED: vicilin	27	5.48E+08	tr|A0A5E4EZP4| A0A5E4EZP4	0.2
tr|A0A5E4FV72|A0A5E4FV72	PREDICTED: vicilin	37	4.63E+08	tr|A0A5E4FV72| A0A5E4FV72	0.2
tr|A0A5E4EE27|A0A5E4EE27	PREDICTED: vicilin	27	2.71E+08	tr|A0A5E4EE27| A0A5E4EE27	0.1
tr|A0A5E4EKE0|A0A5E4EKE0	PREDICTED: peroxygenase	29	3.63E+07	tr|A0A5E4EKE0| A0A5E4EKE0	0.0

Table 8.S13. Identification of EAEP3 band from SDS-PAGE by LC-MS/MS analysis of tryptic peptides produced by in-gel protein digestion.

Protein accession	Protein name	Coverage	Area	Protein accession for coverage and area	Relative abundance (%)

tr|A0A5E4FK23|A0A5E4FK23:tr|E3SH29|E3SH29:tr|A0A4Y1S2I9|A0A4Y1S2I9:tr|Q43608|Q43608	Prunin 2	62	1.70E+11	tr|E3SH29|E3SH29	63.1
tr|A0A5E4FFS0|A0A5E4FFS0:sp|E3SH28|PRU01:sp|Q43607|PRU1	Prunin 1	74	9.14E+10	sp|E3SH28|PRU01	33.9
tr|A0A5E4EZP4|A0A5E4EZP4	PREDICTED: vicilin	35	2.78E+09	tr|A0A5E4EZP4|A0A5E4EZP4	1.0
tr|A0A5E4FV72|A0A5E4FV72	PREDICTED: vicilin	40	2.61E+09	tr|A0A5E4FV72|A0A5E4FV72	1.0
tr|A0A5E4EE27|A0A5E4EE27	PREDICTED: vicilin	44	2.59E+09	tr|A0A5E4EE27|A0A5E4EE27	1.0
tr|A0A5E4E244|A0A5E4E244	PREDICTED: vicilin	18	4.82E+07	tr|A0A5E4E244|A0A5E4E244	0.0
tr|A0A5E4EYT9|A0A5E4EYT9:tr|A0A5E4EYX0|A0A5E4EYX0:tr|A0A516F3L2|A0A516F3L2	Cysteine rich antimicrobial protein	29	2.26E+07	tr|A0A5E4EYT9|A0A5E4EYT9	0.0
tr|A0A5E4EKE0|A0A5E4EKE0	PREDICTED: peroxygenase	21	1.81E+07	tr|A0A5E4EKE0|A0A5E4EKE0	0.0
tr|A0A4Y1RB79|A0A4Y1RB79:tr|A0A5E4FGF6|A0A5E4FGF6:tr|A0A5H2XM09|A0A5H2XM09	Glyceraldehyde-3-phosphate dehydrogenase	16	8.70E+06	tr|A0A4Y1RB79|A0A4Y1RB79	0.0

Table 8.S14. Identification of EAEP4 band from SDS-PAGE by LC-MS/MS analysis of tryptic peptides produced by in-gel protein digestion.

Protein accession	Protein name	Coverage	Area	Protein accession for coverage and area	Relative abundance (%)
tr|A0A5E4FFS0|A0A5E4FFS0:sp|E3SH28|PRU01:sp|Q43607|PRU1	Prunin 1	83	5.11E+10	sp|E3SH28|PRU01	63.7
tr|A0A5E4EZP4|A0A5E4EZP4:tr|A0A4Y1QPK2|A0A4Y1QPK2	PREDICTED: vicilin	52	1.41E+10	tr|A0A5E4EZP4|A0A5E4EZP4	17.5
tr|E3SH29|E3SH29:tr|Q43608|Q43608:tr|A0A4Y1S2I9|A0A4Y1S2I9:tr|A0A5E4FK23|A0A5E4FK23	Prunin 2	65	1.14E+10	tr|E3SH29|E3SH29	14.2
tr|A0A5E4EE27|A0A5E4EE27	PREDICTED: vicilin	53	2.64E+09	tr|A0A5E4EE27|A0A5E4EE27	3.3
tr|A0A5E4EYX0|A0A5E4EYX0:tr|A0A5E4EYT9|A0A5E4EYT9:tr|A0A516F3L2|A0A516F3L2	Cysteine rich antimicrobial protein	42	4.35E+08	tr|A0A5E4EYX0|A0A5E4EYX0	0.5
tr|A0A5E4FV72|A0A5E4FV72	PREDICTED: vicilin	33	2.34E+08	tr|A0A5E4FV72|A0A5E4FV72	0.3
tr|A0A4Y1QPI1|A0A4Y1QPI1:tr|A0A5E4F2T7|A0A5E4F2T7	PREDICTED: vicilin	23	2.25E+08	tr|A0A4Y1QPI1|A0A4Y1QPI1	0.3
tr|A0A5E4ECQ1|A0A5E4ECQ1:tr|A0A5H2Y344|A0A5H2Y344	NAD(P)-binding Rossmann-fold superfamily protein	23	3.15E+07	tr|A0A5E4ECQ1|A0A5E4ECQ1	0.0
tr|A0A5E4GEN6|A0A5E4GEN6:sp|Q945K2|MDL2	(R)-mandelonitrile lyase 2	14	2.34E+07	sp|Q945K2|MDL2	0.0
tr|A0A5E4FH45|A0A5E4FH45:tr|A0A4Y1RM23|A0A4Y1RM23	PREDICTED: putative	53	2.17E+07	tr|A0A5E4FH45|A0A5E4FH45	0.0
tr|A0A5E4EYP9|A0A5E4EYP9	MD-2-related lipid recognition	15	1.85E+07	tr|A0A5E4EYP9|A0A5E4EYP9	0.0

	domain-containing protein / ML domain-containing protein				
sp|Q43804|OLEO1:tr|A0A5E4EAT1|A0A5E4EAT1	Oleosin 1	34	1.43E+07	sp|Q43804|OLEO1	0.0
tr|A0A5E4EKE0|A0A5E4EKE0	PREDICTED: peroxygenase	20	1.06E+07	tr|A0A5E4EKE0|A0A5E4EKE0	0.0
tr|A0A5E4F472|A0A5E4F472:tr|A0A4Y1RDA8|A0A4Y1RDA8	Endoribonuclease L-PSP family protein	34	9.82E+06	tr|A0A5E4F472|A0A5E4F472	0.0

Table 8.S15. Identification of EAEP5 band from SDS-PAGE by LC-MS/MS analysis of tryptic peptides produced by in-gel protein digestion.

Protein accession	Protein name	Coverage	Area	Protein accession for coverage and area	Relative abundance (%)
tr|A0A5E4FFS0|A0A5E4FFS0:sp|E3SH28|PRU01:sp|Q43607|PRU1	Prunin 1	83	1.87E+11	sp|E3SH28|PRU01	74.4
tr|E3SH29|E3SH29:tr|A0A5E4FK23|A0A5E4FK23:tr|A0A4Y1S2I9|A0A4Y1S2I9:tr|Q43608|Q43608	Prunin 2	69	5.72E+10	tr|E3SH29|E3SH29	22.7
tr|A0A5E4EZP4|A0A5E4EZP4:tr|A0A4Y1QPK2|A0A4Y1QPK2	PREDICTED: vicilin	54	6.00E+09	tr|A0A5E4EZP4|A0A5E4EZP4	2.4
tr|A0A5E4EYX0|A0A5E4EYX0:tr|A0A5E4EYT9|A0A5E4EYT9:tr|A0A516F3L2|A0A516F3L2	Cysteine rich antimicrobial protein	56	3.63E+08	tr|A0A5E4EYX0|A0A5E4EYX0	0.1
tr|A0A5E4F2T7|A0A5E4F2T7:tr|A0A4Y1QPI1|A0A4Y1QPI1	PREDICTED: vicilin	24	2.84E+08	tr|A0A4Y1QPI1|A0A4Y1QPI1	0.1
tr|A0A5E4EE27|A0A5E4EE27	PREDICTED: vicilin	45	2.77E+08	tr|A0A5E4EE27|A0A5E4EE27	0.1
tr|A0A5E4FV72|A0A5E4FV72	PREDICTED: vicilin	28	2.17E+08	tr|A0A5E4FV72|A0A5E4FV72	0.1
tr|A0A4Y1RRI6|A0A4Y1RRI6	Non-specific lipid-transfer protein	11	1.17E+08	tr|A0A4Y1RRI6|A0A4Y1RRI6	0.0
tr|A0A5E4ECQ1|A0A5E4ECQ1:tr|A0A5H2Y344|A0A5H2Y344	NAD(P)-binding Rossmann-fold superfamily protein	33	6.06E+07	tr|A0A5E4ECQ1|A0A5E4ECQ1	0.0
tr|A0A5E4EYP9|A0A5E4EYP9	PREDICTED: 11-beta-hydroxysteroid	25	2.25E+07	tr|A0A5E4EYP9|A0A5E4EYP9	0.0
tr|A0A4Y1QWY8|A0A4Y1QWY8	Beta-tonoplast intrinsic protein	16	1.68E+07	tr|A0A4Y1QWY8|A0A4Y1QWY8	0.0
tr|A0A5E4ET55|A0A5E4ET55	Oleosin	9	1.44E+07	tr|A0A5E4ET55|A0A5E4ET55	0.0
tr|A0A5E4FH45|A0A5E4FH45	PREDICTED: putative	60	8.89E+06	tr|A0A5E4FH45|A0A5E4FH45	0.0
tr|A0A4Y1RDA8|A0A4Y1RDA8:tr|A0A5E4F472|A0A5E4F472	Endoribonuclease L-PSP family protein	37	8.39E+06	tr|A0A4Y1RDA8|A0A4Y1RDA8	0.0

tr|A0A5E4E244|A0A5E4E244	PREDICTED: vicilin	10	5.24E+06	tr|A0A5E4E244|A0A5E4E244	0.0	

Table 8.S16. List of almond allergen proteins in online databases[a]

Allergen	Protein name	Isoallergen and variants	MW (kDa)[b]	AAs	UniProt accession	WHO-IUIS
Pru du 1	PR-10		17.6	160	B6CQR7	Yes
	PR-10		17.4	160	B6CQR9	
	PR-10		17.2	160	B6CQS1	
	PR-10		17.7	159	B6CQS3	
	PR-10		17.4	160	B6CQS5	
	PR-10		17.4	160	B6CQS7	
	PR-10		17.4	160	B6CQS9	
Pru du 2	Thaumatin		25.9	246	B6CQT2	No
	Thaumatin		25.8	246	B6CQT4	
	Thaumatin		25.8	246	B6CQT6	
	Thaumatin		30.1	277	B6CQT8	
	Thaumatin		34.2	330	B6CQU0	
Pru du 3	Non-specific lipid-transfer protein 1		11.9	117	Q43017	Yes
	Non-specific lipid-transfer protein		11.9	117	E7CLR2	
	Non-specific lipid-transfer protein 3		12.5	123	Q43019	
	Non-specific lipid-transfer protein	Pru du 3.0101	12.5	123	C0L0I5	
Pru du 4	Profilin		14.0	131	B6CQV0	Yes
	Profilin	Pru du 4.0101, Pru du 4.0102	14.1	131	Q8GSL5	
Pru du 5	60S Acidic ribosomal protein	Pru du 5.0101	11.4	113	Q8H2B9	Yes
Pru du 6	Prunin 1		63.0	551	Q43607	Yes
	Prunin 1	Pru du 6.0101	63.1	551	E3SH28	
	Prunin 2	Pru du 6.0201	57.0	504	E3SH29	
	Prunin 2		57.1	504	Q43608	
Pru du 8	Cysteine rich antimicrobial protein	Pru du 8.0101	31.1	264	A0A516F3L2	Yes
Pru du 10	Mandelonitrile lyase 2		61.2	563	Q945K2	Yes
Pru du AP	γ-Conglutin 1		46.9	431	P82952 (A0A5E4EAH6)	No

[a] Allergen information was obtained from the allergen databases of the World Health Organization and the International Union of Immunological Societies (WHO/IUIS) Allergen Nomenclature Sub-committee (http://www.allergen.org/, accessed 2/15/2022) and Allergome (http://www.allergome.org/, accessed 2/15/2022).

[b] Molecular weight (MW) and number of amino acids (AAs) represent the information for unprocessed proteins.

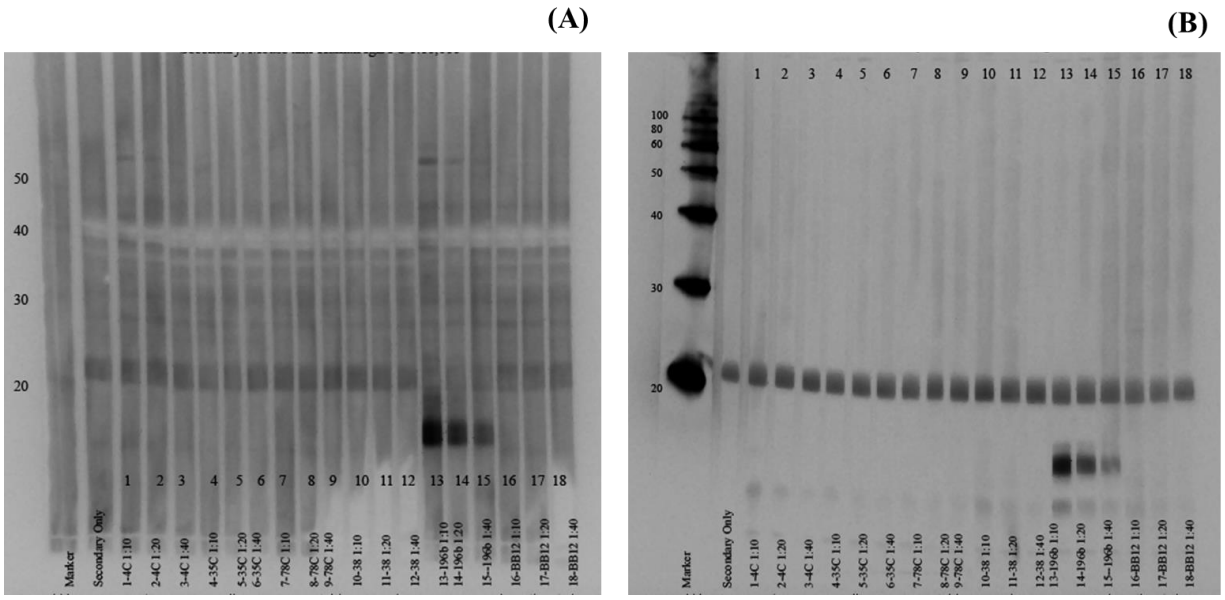


Fig. 8.S1. Western blots of AEP (A) and EAEP (B) samples. Primary human sera: 1:10, 1:20, 140; Secondary: Mouse anti-Human IgE Fc: 1:10,000.

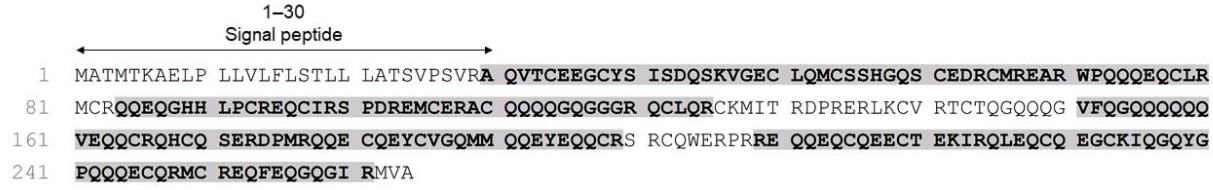


Fig. 8.S2. Sequence coverage of cysteine-rich antimicrobial protein (UniProt accession A0A516F3L2; Pru du 8) in AEP 9 gel cut. Amino acids included in the identified tryptic peptide sequences in proteomics analysis are shaded gray.

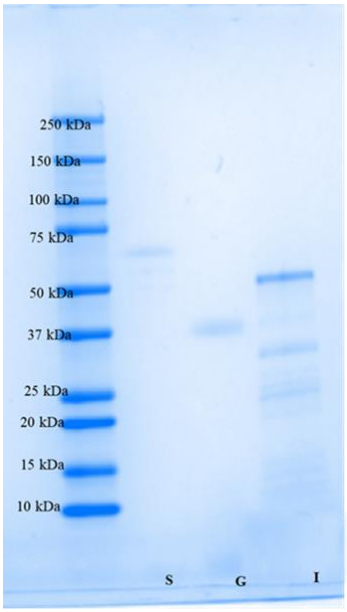


Fig. 8.S3. SDS-PAGE of simulated saliva (S) simulated gastric (G) and simulated intestinal (I) fluids.

References

Ahrens, S., Venkatachalam, M., Mistry, A. M., Lapsley, K., & Sathe, S. K. (2005). Almond (Prunus dulcis L.) protein quality. *Plant Foods for Human Nutrition*, *60*(3), 123–128. https://doi.org/10.1007/s11130-005-6840-2

Akharume, F. U., Aluko, R. E., & Adedeji, A. A. (2021). Modification of plant proteins for improved functionality: A review. *Comprehensive Reviews in Food Science and Food Safety*, *20*(1), 1541-4337.12688. https://doi.org/10.1111/1541-4337.12688

Almeida, N. M. de, Dias, F. F. G., Rodrigues, M. I., & Bell, J. M. L. N. de M. (2019). Effects of Processing Conditions on the Simultaneous Extraction and Distribution of Oil and Protein from Almond Flour. *Processes 2019, Vol. 7, Page 844*, *7*(11), 844. https://doi.org/10.3390/PR7110844

Alonso, R., Aguirre, A., & Marzo, F. (2000). Effects of extrusion and traditional processing methods on antinutrients and in vitro digestibility of protein and starch in faba and kidney beans. *Food Chemistry*, *68*(2), 159–165. https://doi.org/10.1016/S0308-8146(99)00169-7

Amirshaghaghi, Z., Rezaei, K., & Rezaei, M. H. (2017). Characterization and functional properties of protein isolates from wild almond. *Journal of Food Measurement and Characterization*, *11*(4), 1725–1733. https://doi.org/10.1007/s11694-017-9553-y

Bornhorst, G. M., & Singh, R. P. (2013). Kinetics of in Vitro Bread Bolus Digestion with Varying Oral and Gastric Digestion Parameters. *Food Biophysics*, *8*(1), 50–59. https://doi.org/10.1007/s11483-013-9283-6

Cabanillas, B., Pedrosa, M. M., Rodríguez, J., González, Á., Muzquiz, M., Cuadrado, C., Crespo, J. F., & Burbano, C. (2010a). Effects of enzymatic hydrolysis on lentil allergenicity. *Molecular Nutrition and Food Research*, *54*(9), 1266–1272. https://doi.org/10.1002/mnfr.200900249

Cabanillas, B., Pedrosa, M. M., Rodríguez, J., Muzquiz, M., Maleki, S. J., Cuadrado, C., Burbano, C., & Crespo, J. F. (2012). Influence of Enzymatic Hydrolysis on the Allergenicity of Roasted Peanut Protein Extract. *International Archives of Allergy and Immunology*, *157*(1), 41–50. https://doi.org/10.1159/000324681

Che, H., Zhang, Y., Jiang, S., Jin, T., Lyu, S. C., Nadeau, K. C., & McHugh, T. (2019). Almond (Prunus dulcis) Allergen Pru du 8, the First Member of a New Family of Food Allergens. *Journal of Agricultural and Food Chemistry*, *67*(31), 8626–8631. https://doi.org/10.1021/ACS.JAFC.9B02781/SUPPL_FILE/JF9B02781_SI_001.PDF

Clemente, A., Vioque, J., Sanchez-Vioque, R., Pedroche, J., & Millán, F. (1999). Production of Extensive Chickpea (Cicer arietinum L.) Protein Hydrolysates with Reduced Antigenic Activity. *Journal of Agricultural and Food Chemistry*, *47*(9), 3776–3781. https://doi.org/10.1021/JF981315P

de Souza, T. S. P., Dias, F. F. G., Oliveira, J. P. S., de Moura Bell, J. M. L. N., & Koblitz, M. G. B. (2020a). Biological properties of almond proteins produced by aqueous and enzyme-assisted aqueous extraction processes from almond cake. *Scientific Reports*, *10*(1), 1–12. https://doi.org/10.1038/s41598-020-67682-3

de Souza, T. S. P., Dias, F. F. G., Oliveira, J. P. S., de Moura Bell, J. M. L. N., & Koblitz, M. G. B. (2020b). Biological properties of almond proteins produced by aqueous and enzyme-assisted aqueous extraction processes from almond cake. *Scientific Reports*, *10*(1), 10873. https://doi.org/10.1038/s41598-020-67682-3

Dias, F. F. G., Almeida, N. M., Souza, T. S. P., Taha, A. Y., & Bell, J. M. L. N. M. (2020). Characterization and Demulsification of the Oil-Rich Emulsion from the Aqueous Extraction Process of Almond Flour. *Processes*, *8*(10), 1228. https://doi.org/10.3390/pr8101228

Dias, F. F. G., & de Moura Bell, J. M. L. N. (2022). Understanding the impact of enzyme-assisted aqueous extraction on the structural, physicochemical, and functional properties of protein extracts from full-fat almond flour. *Food Hydrocolloids*, *127*, 107534. https://doi.org/10.1016/j.foodhyd.2022.107534

Dias, F. F. G., Taha, A. Y., & Bell, L. N. de M. (2022). Effects of enzymatic extraction on the simultaneous extraction of oil and protein from full-fat almond flour, insoluble microstructure, emulsion stability and functionality. *Future Foods*, *5*, 100151. https://doi.org/10.1016/J.FUFO.2022.100151

Dreveny, I., Gruber, K., Glieder, A., Thompson, A., & Kratky, C. (2001). The Hydroxynitrile Lyase from Almond: A Lyase that Looks Like an Oxidoreductase. *Structure*, *9*(9), 803–815. https://doi.org/10.1016/S0969-2126(01)00639-6

El-Aal, M. H. A., Hamza, M. A., & Rahma, E. H. (1986). In vitro digestibility, physico-chemical and functional properties of apricot kernel proteins. *Food Chemistry*, *19*(3), 197–211. https://doi.org/10.1016/0308-8146(86)90070-1

Garcia-Mas, J., Messeguer, R., Arús, P., & Puigdomènech, P. (1995). Molecular characterization of cDNAs corresponding to genes expressed during almond (Prunus amygdalus Batsch) seed development. *Plant Molecular Biology*, *27*(1), 205–210. https://doi.org/10.1007/BF00019192

Geiselhart, S., Hoffmann-Sommergruber, K., & Bublin, M. (2018). Tree nut allergens. *Molecular Immunology*. https://doi.org/10.1016/j.molimm.2018.03.011

Grimble, G. K. (1991). The significance of peptides in clinical nutrition. *Annu. Rev. Nutr*, *10*(SUPPL.), 419–447. https://doi.org/10.1016/0261-5614(91)90110-X

Gundry, R. L., White, M. Y., Murray, C. I., Kane, L. A., Fu, Q., Stanley, B. A., & Van Eyk, J. E. (2009). Preparation of proteins and peptides for mass spectrometry analysis in a bottom-up proteomics workflow. *Current Protocols in Molecular Biology*, *SUPPL. 88*, 1–23. https://doi.org/10.1002/0471142727.mb1025s88

Gupta, R. S., Warren, C. M., Smith, B. M., Jiang, J., Blumenstock, J. A., Davis, M. M., Schleimer, R. P., & Nadeau, K. C. (2019). Prevalence and Severity of Food Allergies Among US Adults. *JAMA Network Open*, *2*(1), e185630–e185630. https://doi.org/10.1001/JAMANETWORKOPEN.2018.5630

HARRY, TOWBIN, THEOPHIL STAEHELINt, A. J. G. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: Procedure and some applications. *Proc. Natl. Acad. Sci. USA*, *79*, 4350-4354,.

He, S., Simpson, B. K., Ngadi, M. O., & Ma, Y. (2015). In vitro studies of the digestibility of lectin from black turtle bean (Phaseolus vulgaris). *Food Chemistry*, *173*, 397–404. https://doi.org/10.1016/j.foodchem.2014.10.045

Huang, Y. P., Dias, F. F. G., Leite Nobrega de Moura Bell, J. M., & Barile, D. (2022). A complete workflow for discovering small bioactive peptides in foods by LC-MS/MS: A case study on almonds. *Food Chemistry*, *369*, 130834. https://doi.org/10.1016/j.foodchem.2021.130834

Kabasser, S., Hafner, C., Chinthrajah, S., Sindher, S. B., Kumar, D., Kost, L. E., Long, A. J., Nadeau, K. C., Breiteneder, H., & Bublin, M. (2021). Identification of Pru du 6 as a potential marker allergen for almond allergy. *Allergy*, *76*(5), 1463–1472. https://doi.org/10.1111/ALL.14613

Konieczny, D., Stone, A. K., Nosworthy, M. G., House, J. D., Korber, D. R., Nickerson, M. T., & Tanaka, T. (2020). Nutritional properties of pea protein-enriched flour treated with different proteases to varying degrees of hydrolysis. *Cereal Chemistry*, *97*(2), 429–440. https://doi.org/10.1002/CCHE.10258

Laemmli, U. K. (1970). *Cleavage of structural proteins during the assembly of the head of bacteriophage {T}4*. *227*, 680–685. https://doi.org/10.1038/227680a0

Nielsen, P.M., Petersen, D. and Dambmann, C. (2001). Improved Method for Determining. *Food Chemistry*, *66*(5), 642–646. https://doi.org/10.1109/TPAMI.2012.61

Poltronieri, P., Cappello, M. S., Dohmae, N., Conti, A., Fortunato, D., Pastorello, E. A., Ortolani, C., & Zacheo, G. (2002). Identification and characterisation of the IgE-binding proteins 2S albumin and conglutin γ in almond (Prunus dulcis) seeds. *International Archives of Allergy and Immunology*, *128*(2), 97–104. https://doi.org/10.1159/000059399

Sathe, S. K. (1993). Solubilization, electrophoretic characterization and in vitro digestibility of almond (Prunus amygdalus) proteins. *Journal of Food Biochemistry*, *16*, 249–264.

Sathe, S. K., Wolf, W. J., Roux, K. H., Teuber, S. S., Venkatachalam, M., & Sze-Tao, K. W. C. (2002). Biochemical Characterization} of Amandin, the Major Storage Protein in Almond (Prunus dulcis L.). *J. Agric. Food Chem.*, *50*(15), 4333–4341. https://doi.org/10.1021/jf020007v

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of Image Analysis. *Nature Methods*, *9*(7), 671. https://doi.org/10.1038/NMETH.2089

Sicherer, S. H., Muñoz-Furlong, A., & Sampson, H. A. (2003). Prevalence of peanut and tree nut allergy in the United States determined by means of a random digit dial telephone survey: a 5-year follow-up study. *The Journal of Allergy and Clinical Immunology*, *112*(6), 1203–1207. https://doi.org/10.1016/S0091-6749(03)02026-8

Souza, T. S. P., Dias, F. F. G., Koblitz, M. G. B., & Bell, J. M. L. N. M. (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

Sze-Tao, K. W. C., & Sathe, S. K. (2000). Functional properties and in vitro digestibility of almond (Prunus dulcis L.) protein isolate. *Food Chemistry*, *69*(2), 153–160. https://doi.org/10.1016/S0308-8146(99)00244-7

Tiwari, R. S., Venkatachalam, M., Sharma, G. M., Su, M., Roux, K. H., & Sathe, S. K. (2010). LWT - Food Science and Technology Effect of food matrix on amandin , almond (Prunus dulcis L .) major protein , immunorecognition and recovery. *LWT - Food Science and Technology*, *43*(4), 675–683. https://doi.org/10.1016/j.lwt.2009.11.012

Tomishima, H., Luo, K., & Mitchell, A. E. (2021). *The Almond (Prunus dulcis): Chemical Properties, Utilization, and Valorization of Coproducts*. https://doi.org/10.1146/annurev-food-052720

Tree Nuts: World Markets and Trade. (2021). *USDA*. https://public.govdelivery.com/accounts/USDAFAS/subscriber/new

Verhoeckx, K. C. M., Vissers, Y. M., Baumert, J. L., Faludi, R., Feys, M., Flanagan, S., Herouet-Guicheney, C., Holzhauser, T., Shimojo, R., van der Bolt, N., Wichers, H., & Kimber, I. (2015). Food processing and allergenicity. *Food and Chemical Toxicology*, *80*, 223–240. https://doi.org/10.1016/j.fct.2015.03.005

Willison, L. A. N., Tripathi, P., Sharma, G., Teuber, S. S., Sathe, S. K., & Roux, K. H. (2011). Cloning, Expression and Patient IgE Reactivity of Recombinant Pru du 6, an 11S Globulin from Almond. *International Archives of Allergy and Immunology*, *156*(3), 267–281. https://doi.org/10.1159/000323887

Wolf, W. J., & Sathe, S. K. (1998). Ultracentrifugal and polyacrylamide gel electrophoretic studies of extractability and stability of almond meal proteins. *Journal of the Science of Food and Agriculture*, *78*(4), 511–521. https://doi.org/10.1002/(SICI)1097-0010(199812)78:4<511::AID-JSFA148>3.0.CO;2-X

Yada, S., Lapsley, K., & Huang, G. (2011). A review of composition studies of cultivated almonds: {Macronutrients} and micronutrients. *Journal of Food Composition and Analysis*, *24*(4–5), 469–480. https://doi.org/10.1016/j.jfca.2011.01.007