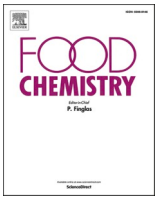
Food Chemistry 409 (2023) 135267

Contents lists available at ScienceDirect

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem



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

Yu-Ping Huang [a], Bruna Paviani [a], Naomi K. Fukagawa [b], Katherine M. Phillips [c], Daniela Barile [a,d,*]

[a] Department of Food Science and Technology, University of California, Davis, One Shields Avenue, Davis, CA 95616, United States
[b] USDA ARS Beltsville Human Nutrition Research Center, 10300 Baltimore Ave, BARC-East, Center Road, Beltsville, MD 20705, United States
[c] Department of Biochemistry, Virginia Tech, Blacksburg, VA 24061, United States
[d] Foods for Health Institute, University of California, Davis, One Shields Avenue, Davis, CA 95616, United States

ARTICLE INFO

Keywords:
α-Galactooligosaccharides
HPAE-PAD
LC-Q-TOF MS
2,3-Butanediol glycosides
Low molecular weight soluble dietary fiber
Soybeans

ABSTRACT

Oligosaccharides are known for several bioactivities on health, however, in sensitive individuals, can cause intestinal discomfort. This study aimed to investigate the oligosaccharide profiles in selected plant-based food products. A quantification method based on high-performance anion-exchange chromatography-pulsed amperometric detection was developed, validated, and used to measure major oligosaccharides. Additional low-abundant oligosaccharides and glycosides were characterized by liquid chromatography-tandem mass spectrometry and glycosidases. The summed concentration of raffinose, stachyose, and verbascose ranged from 0.12–0.19 mg/g in almond milk, 3.6–6.4 mg/g in soy milk, and 74–77 and 4.8–57 mg/g in defatted and full-fat soy four. Over 80 different oligosaccharides were characterized. Novel compounds, 2,3-butanediol glycosides, were identified in almond milk. Low-abundant oligosaccharides represented 25 %, 6 %, and 10 % of total OS in almond milk, soy milk, and soy flour, respectively. The data here are useful to estimate oligosaccharide consumption from dietary intake and facilitate further studies on their bioactivity.

1. Introduction

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 resulted in increased development and consumption of plant-based beverages, such as almond milk, soy milk, and others, as alternatives to cow's milk. Almond milk is currently the most popular milk alternative in the United States (Wunsch, 2022). Although traditionally widely consumed in several Asian countries, soy milk has only recently become more prevalent in Western countries. 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.

Soybean and almond seeds both contain naturally occurring oligosaccharides (OS), predominated by α-galactooligosaccharides, including raffinose, stachyose, and verbascose. The consumption of soybean products may cause flatulence due to the high abundance of α-galactooligosaccharides in soybeans (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, α-galactooligosaccharides have been shown to exhibit potential prebiotic activity, such as reshaping bacterial composition by increasing the relative abundance of *Bifidobacterium* and *Lactobacillus* 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 et al., 2017; Dai et al., 2019) and reducing appetite and inflammation in overweight adults (Morel et al., 2015).

Although a few studies report the contents of some OS in almonds and soybeans (Barreira et al., 2010; Fan et al., 2015; Kuo et al., 1988),

* Corresponding author at: Department of Food Science and Technology, University of California, Davis, One Shields Avenue, Davis, CA 95616, United States.
E-mail address: dbarile@ucdavis.edu (D. Barile).

https://doi.org/10.1016/j.foodchem.2022.135267
Received 29 June 2022; Received in revised form 16 December 2022; Accepted 19 December 2022
Available online 21 December 2022

those values cannot be used for extrapolating the OS concentration in commercial almond and soy products. Besides the variation in plant materials due to varieties, geography, and growing conditions (Bainy et al., 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.

In addition to α-galactooligosaccharides, other OS present in plant-based foods may be involved in gut microbial fermentation and should be included in the characterization. Liquid chromatography-tandem mass spectrometry (LC-MS/MS) can be used to comprehensively characterize OS through deducing structural information based on fragmentation patterns. Recently, our group reported the optimization of a conventional glycomics method to improve the identification of low-abundant OS (Huang, Robinson, Dias, et al., 2022) and avoid incorrect identification caused by unexpected oligosaccharide degradation (Huang, Robinson, & Barile, 2022), enabling its application to the identification of plant OS.

In this study, a quantification method based on high-performance anion-exchange chromatography with pulsed amperometric detection (HPAE-PAD) was established and used to measure major OS in selected soy and almond products. Furthermore, unexplored OS and glycosides were characterized by LC-MS/MS and an enzymatic approach with selected glycosidases.

2. Materials and methods

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), obtained from local stores or purchased online in the United States (Supplementary Material Table 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, where information about other food components, including proximate composition and others, may be found (FoodData Central, 2022). Two additional soy flour samples, NIST SRM® 3234 (National Institute of Standards and Technology, Gaithersburg, MD, USA) and in-house full-fat soy flour control material (Soy Flour CC), were analyzed in each batch for quality control. Additional samples, including an almond milk product and a soy milk product were obtained from local stores (Davis, CA, USA) and a full-fat soy flour product was purchased online in the United States and used for optimizing the oligosaccharide extraction method.

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).

2.3. Soy flour defatting

n-Hexane was added to 20–25 mg of full-fat soy flour samples (solvent-to-material ratio 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.

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

2.4.1. Method development for extraction of OS from almond milk and soy milk

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 1000 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.

2.4.1.2. Ethanol precipitation. Protein precipitation using two (2 V) and four (4 V) volumes of ethanol were tested. For the 2 V 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 2 V and 4 V 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 2 V and 4 V 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.

2.4.2. Method development for extraction of OS from soy flour

OS 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 1500 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 1500 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 1500 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 OS by shaking at 1500 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 1500 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 1500 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.

2.4.3. HPAE-PAD analysis of raffinose, stachyose, and verbascose

Quantification of raffinose, stachyose, and verbascose was conducted on a ThermoFisher Dionex ICS-5000 + 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 mM 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).

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-to-noise ratios (S/N) of 3 and 10, respectively.

2.4.5. Method validation to assess the 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 OS 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 OS 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.

2.4.6. Statistical analysis for assessing extraction method efficiency

Extraction efficiencies of different conditions or procedures were compared by one-way analysis of variance (ANOVA) and Tukey's test at $p < 0.05$ in RStudio (3.5.3).

2.4.7. 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 (Supplementary Material Table S1). Samples in the same assay batch were extracted and analyzed on the instrument together.

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 1400 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 1400 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 polyethersulfone (PES) filters and analyzed by HPAE-PAD.

2.5. Comprehensive identification of OS with LC-MS/MS

2.5.1. Oligosaccharide purification prior to liquid chromatography-quadrupole-time-of-flight mass spectrometery (LC-Q-TOF MS) analysis

Almond milk (8 samples, AM1–AM8), soy milk (8 samples SM1–SM8), and 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. OS in soy flour were extracted with water (w/v = 1:20) by shaking at 1400 rpm on a thermomixer at room temperature for 10 min after defatting (full-fat soy flour only). Supernatants were collected after centrifugation at 13,000*g* at room temperature for 15 min. OS were extracted from the liquid samples (i.e., almond milk, soy milk, and the supernatants obtained from soy flour) through ethanol precipitation followed by two stages of solid-phase extraction (SPE) wtih Strata-X-C cartridges (30 mg, Phenomenex, Torrance, CA, USA) and porous graphitic carbon (PGC) SPE microplate with 40 μL chromatographic media bed (Glygen Corporation, Columbia, MD, USA) as previously described (Huang, Robinson, Dias, et al., 2022; Machida et al., 2022). Neutral (fraction 1) and acidic (fraction 2) OS were collected for further analysis (Machida et al., 2022).

2.5.2. LC-Q-TOF MS analysis

The purified OS 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 OS 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 OS 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 OS divided by the total peak area of all the identified OS.

2.6. Identification of selected unknown glycosides

2.6.1. Enzymatic treatment using glucosidases

To identify the unknown glycosides present in almond milk, selected glycosidases were applied to the purified almond oligosaccharide sample. Five microliters of the purified almond neutral oligosaccharide 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 OS 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 OS 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.

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 to 2 min and increased to 45 % B from 2 to 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.

3. Results and discussion

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

3.1.1. HPAE-PAD methodology

An HPAE-PAD method using a CarboPac PA200 column with isocratic elution was developed to separate and quantify major OS in the almond and soy samples. Fig. 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 et al., 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 OS analyzed. Therefore, quadratic calibration curves were chosen to determine the oligosaccharide concentrations.

3.1.2. Method development for OS extraction

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 OS spiked in the samples.

3.1.2.1. Extraction of almond milk and soy milk. Almond milk and soy milk contain varied protein content (0.44–0.69 g/100 g and 3–4.69 g/100 g, 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 Supplementary Material Table S2. For almond milk, the measured quantities of the OS 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 OS. Bouchard, Hofland, and Witkamp (2007) reported that the solubility of raffinose at

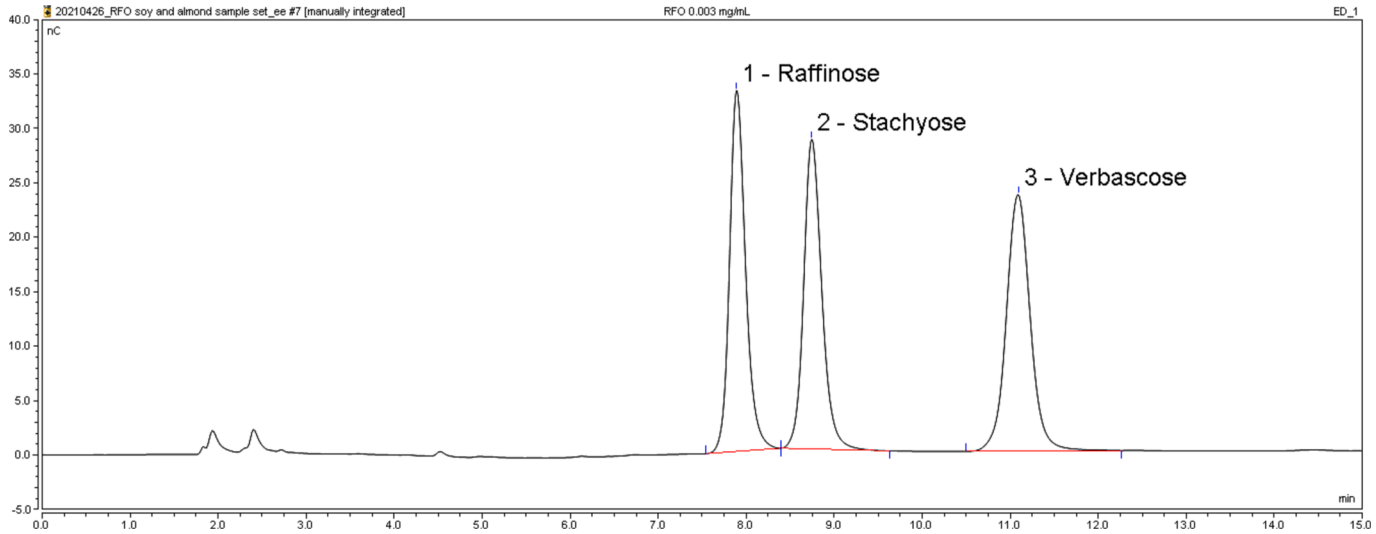


Fig. 1. HPAE-PAD chromatogram of standards of raffinose, stachyose, and verbascose.

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 OS 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 to 107 % (Table 1).

3.1.2.2. 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 OS. 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 Supplementary Material Table 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 OS, was achieved (100, 103, and 103 % for raffinose, stachyose, and verbascose, respectively).

3.1.3. Quantification of raffinose, stachyose, and verbascose in commercial products

Fig. 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

Table 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.

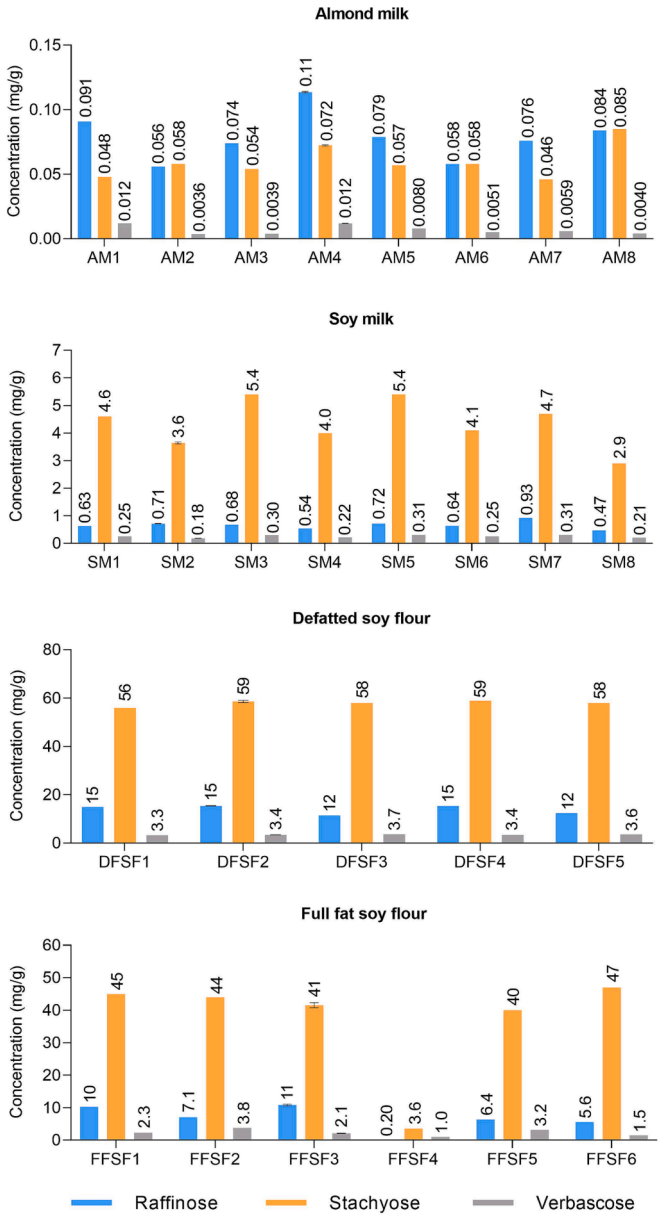


Fig. 2. Concentrations of raffinose, stachyose, and verbascose in commercial almond milk (AM1–AM8), soy milk (SM1–SM8), and soy flour (DFSF1–DFSF5 (defatted) and FFSF1–FFSF6 (full-fat)) products measured by HPAE-PAD. Data of AM4 (n = 3), SM2 (n = 4), DFSF2 (n = 2), and FFSF3 (n = 2) were expressed as mean ± standard deviation to show the measurement uncertainty; data of the other samples (n = 1) represent the measured values.

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; OS 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 OS 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 OS 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 (7.1–21.1 mg/g raffinose) contained lower amounts of OS 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^{-1} 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^{-1}, respectively) (Pico et al., 2021); the instrumental LOQ were 4.9, 6.3, and 12 μg L^{-1}, respectively.

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 OS for which commercial standards are not yet available. To encompass the variety of OS 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 OS besides the three major OS (Supplementary Material Table S4). A total of 82, 60, 48, and 75 OS were identified from the pooled almond milk, soy milk, defatted-, and full-fat soy flour samples, respectively. Most of the OS 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 OS (47, 40, 29, and 51, respectively) comprised only hexoses, with a degree of polymerization of 2–8. Five OS containing pentose units (Hex_4Pent_1, Hex_5Pent_1, and Hex_3Pent_2) were identified (three confirmed by tandem MS, data not shown) in the pooled full-fat soy flour.

Some OS 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. 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 OS were found in all four pooled samples. The three pooled soy samples all contained ten phosphorylated OS (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 OS (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 OS 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. 3A) supported their identifications. Acetylated OS were found in all the soy and almond samples. Five acetylated OS 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 (data not shown). Three of them were also found in common in the almond milk sample.

A series of OS with the *m/z* of 434.187, 596.240, and 758.292 were tentatively identified as glycerol-containing OS. The fragment ions *m/z* 93.05 ($[glycerol + H]^+$), *m/z* 205.107 ($[Hex_1Glycerol_1 + H]^+$), *m/z*

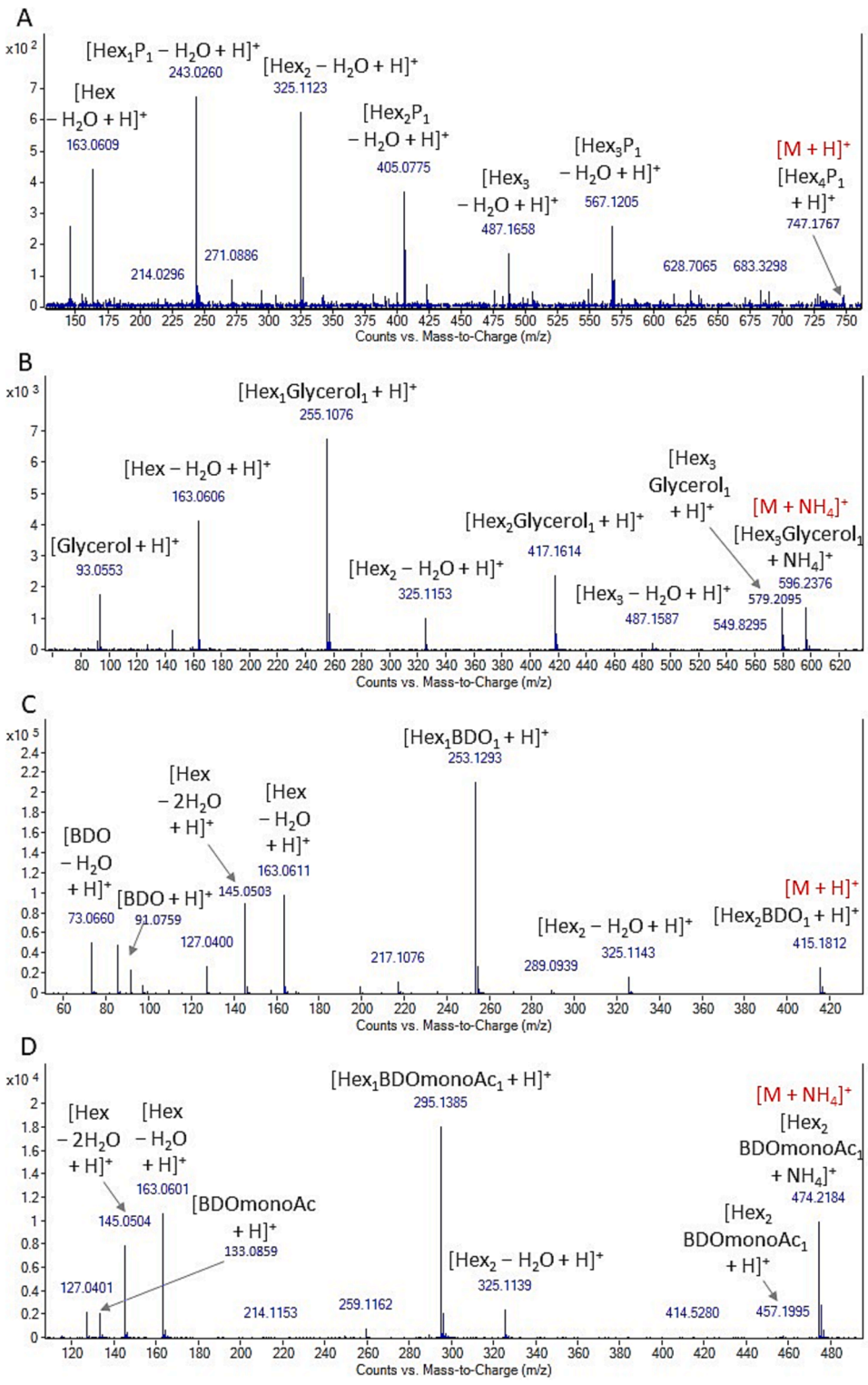


Fig. 3. LC-Q-TOF tandem MS spectra of Hex_4P_1 (13.349 min; spectrum of soy milk; A), Hex_3-$Glycerol_1$ (10.002 min; spectrum of full-fat soy flour; B), and Hex_2BDO_1 (16.623 min; C) and $Hex_2BDOmonoAc_1$ l (18.918 min; D) in almond milk generated from precursor ions of $[M + H]^+$ or $[M + NH_4]^+$ (annotated in red). BDO: butanediol; BDOmonoAc: butanediol monoacetate. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

417.160 ($[Hex_2Glycerol_1 + H]^+$), and *m/z* 579.213 ($[Hex_3Glycerol_1 + H]^+$) in the tandem MS spectra (Fig. 3B) of this series of OS indicates that glycerol constitutes part of the OS. The matched retention times of the OS 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 OS. Glycerol-containing OS were previously identified in wheat flour and algae (Carter et al., 1956; Karsten et al., 2005). Glyceryl glycosides with only one monosaccharide conjugated with glycerol were also found in algae, wine, and sake (Eggert & Karsten, 2010; Ruiz-Matute et al., 2009). However, this type of compounds had not been previously reported in soy and almond products.

Noteworthy, several OS 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 OS 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. 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. 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 et al., 1967) and muskmelon (Lignou et al., 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 et al., 1998) and *in vitro* fecal fermentation product of black rice (Owolabi et al., 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. 4 shows the apparent relative abundance of various OS identified by LC-Q-TOF MS. Interestingly, minor OS 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 ($Hex_{2\text{-}3}$ + butanediol) was the highest class (19 %), followed by $Hex_{3\text{-}8}$ (5 %). Among the three pooled soy samples, OS distribution was similar. Minor OS accounted for 6–10 %, with $Hex_{3\text{-}6/7}$ as the most represented, followed by ciceritol.

3.3. Identity confirmation of 2,3-butanediol glycosides

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. 5A), whereas those same peaks were not

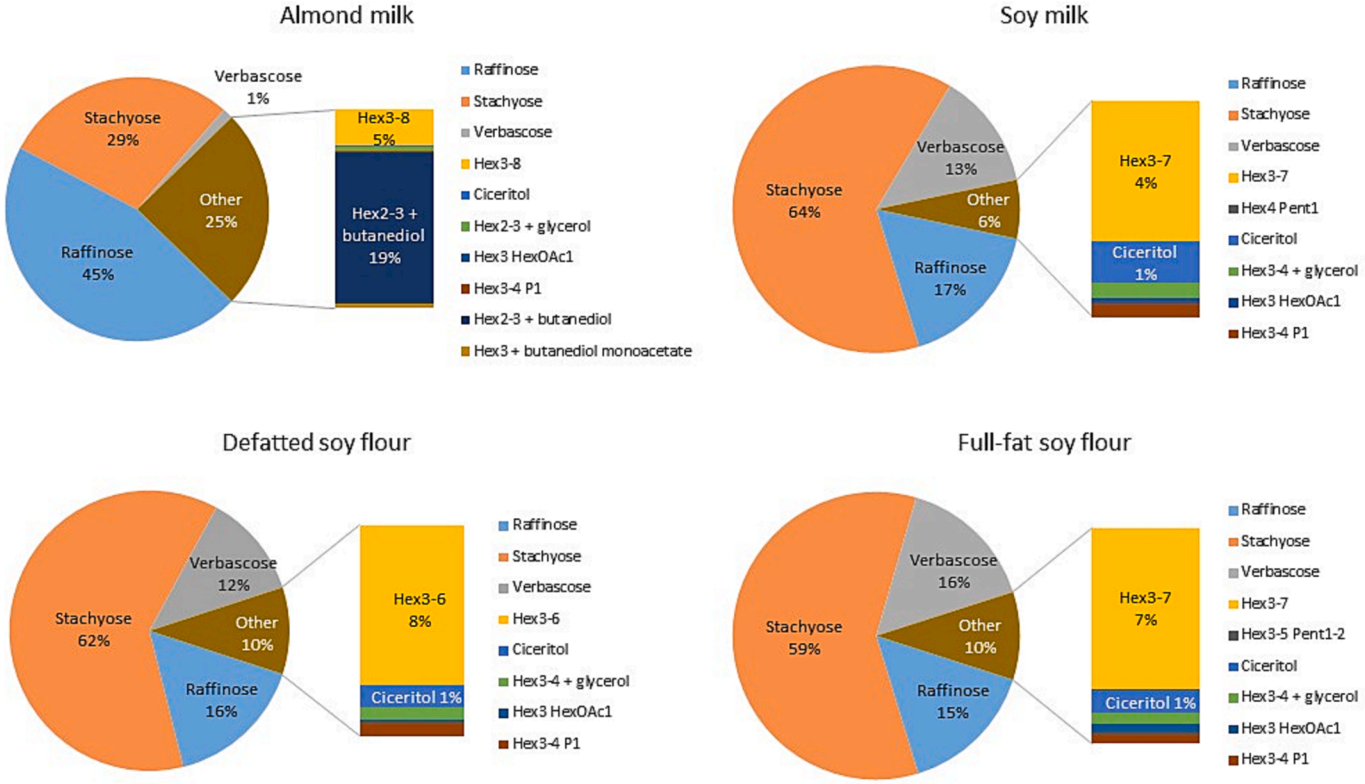


Fig. 4. Apparent relative abundance of different classes of OS identified in the almond milk, soy milk, and soy flour estimated by peak areas from the LC-Q-TOF analysis. $Hex_{3\text{-}6/7/8}$ group excludes raffinose, stachyose, and verbascose. Hex: hexose; Pent: pentose; HexOAc: acetyl-hexose; P: phosphorylation.

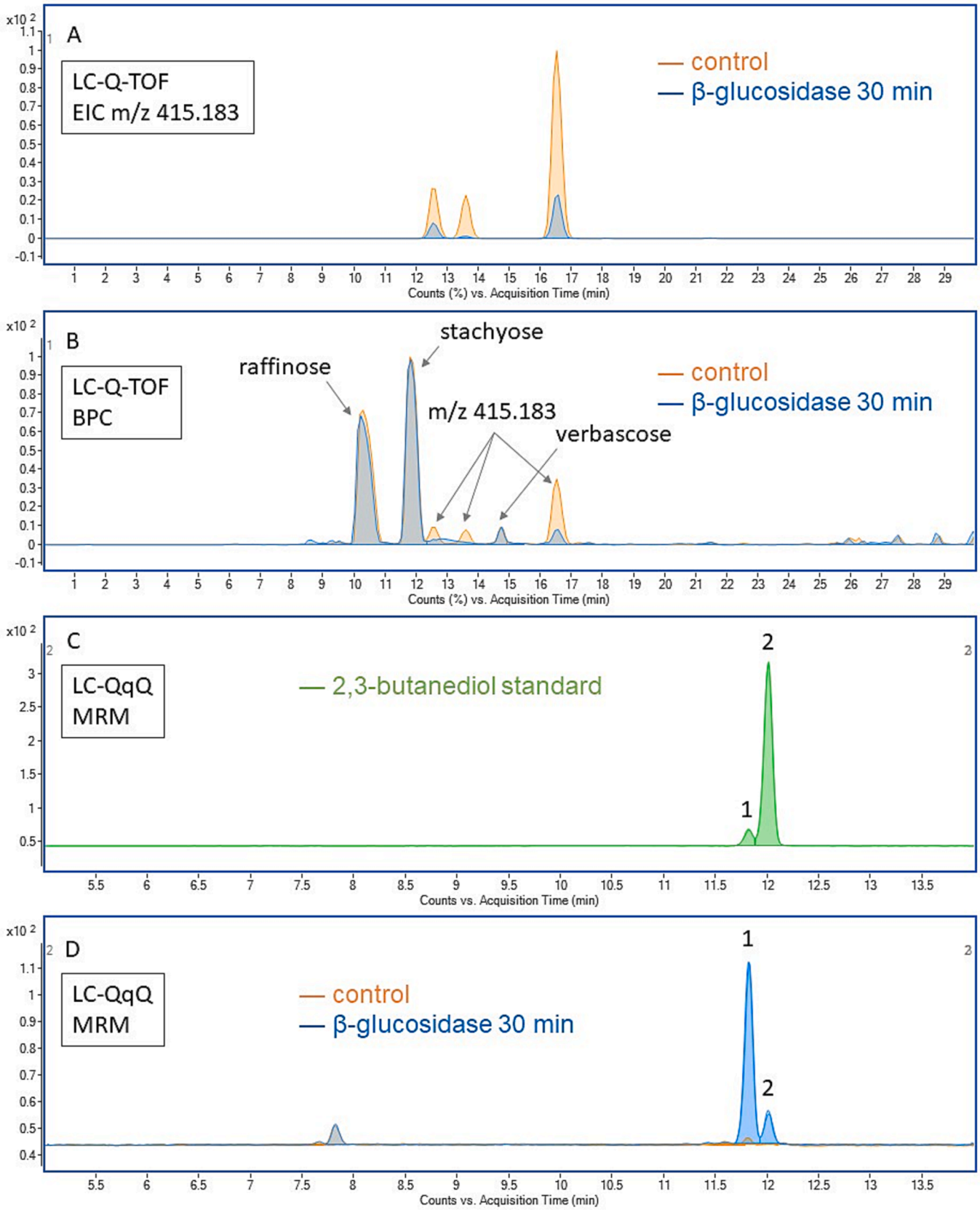


Fig. 5. Overlaid LC-Q-TOF chromatogram of purified almond neutral OS undergoing β-glucosidase treatment (A, extracted ion chromatogram (EIC) of *m/z* 415.183; B, base peak chromatogram (BPC)) showing that the *m/z* 415.183 ion ($[M + H]^+$) peaks decreased after the treatment. Overlaid LC-QqQ MRM chromatograms (*m/z* 484 → 260 and *m/z* 484 → 262) of 2,3-butanediol standard (C) and almond neutral OS with β-glucosidase treatment (D).

affected by α-glucosidase (Supplementary Material Fig. S1A and B). Also, as expected, the peak areas of the major OS, including raffinose, stachyose, and verbascose (containing α-galactosyl and α-1,β-2-glycosidic linkages), were not altered after the enzymatic treatment (Fig. 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 (Supplementary Material Fig. S1C), clearly indicating that at least one terminal hexose unit in this compound was a β-D-glucose.

3.3.2. Analysis of the enzymatically released aglycone by LC-QqQ MS

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. 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. 5D), confirming that β-glucosidase released the two peaks of 2,3-butanediol from the purified almond OS.

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 OS. 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.

4. Conclusions

Non-digestible OS could beneficially affect human health thanks 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 in almond milk, soy milk, and soy flour using HPAE-PAD. The extraction of OS 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 of different brands. The additional analysis by LC-Q-TOF MS allowed the identification of many OS 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 results are novel and extend information about components in food reported in the USDA FoodData Central. Information on the newly identified OS and glycosides can serve as a basis for further investigation into their bioactivity, such as the prebiotic property, to understand their influence to human health.

CRediT authorship 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 M. Phillips:** Conceptualization, Methodology, Writing – review & editing. **Daniela Barile:** Conceptualization, Funding acquisition, Project administration, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The authors thank the U.S. Department of Agriculture (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 USDA Agricultural Research Service (USDA-ARS) (NACA #58-8040-8-013). The graphical abstract was created with BioRender.com (https://biorender.com).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodchem.2022.135267.

References

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*, Article 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

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.

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*, Article 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*, Article 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

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

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