



Protein profile of commercial soybean milks analyzed by label-free quantitative proteomics

Ilaria Battisti^{a,c,1}, Leonard Barnabas Ebinezer^{b,1}, Giovanna Lomolino^b, Antonio Masi^{b,*}, Giorgio Arrigoni^{a,c,d,*}

^a Department of Biomedical Sciences, University of Padova, via U. Bassi 58/B, 35131 Padova, Italy

^b Department of Agronomy, Food, Natural Resources, Animals, and Environment, University of Padova, Via dell'Università 16, 35020 Legnaro, Italy

^c Proteomics Center, University of Padova and Azienda Ospedaliera di Padova, via G. Orus 2/B, 35129 Padova, Italy

^d CRIBI Biotechnology Center, University of Padova, via U. Bassi 58/B, 35131 Padova, Italy

ARTICLE INFO

Keywords:

Plant-based milk
Soy milk
Soy drink
Proteomics
Allergen
Functional drink/beverage

ABSTRACT

The consumption of soy milk is increasing worldwide for its nutritional value and health benefits, however, its protein composition after commercialization is not well known. Technological and thermal treatments to which soy milk is subjected could affect the protein composition of the commercial products. This study compared the protein profile of 15 different commercial soy milks using a label-free quantitative proteomics approach. Proteins related to nutrient reservoir activity, endopeptidase inhibitor activity, lipid binding, and seed maturation contribute the most in terms of percentage mass. Their associated Gene Ontology terms are also enriched. Samples clustered into three groups based on their protein composition, with glycinins and beta-conglycinins being the most influential for determining the clustering. Amino acid composition estimated from the proteomics data also reflects the clustering of samples. Twenty allergenic proteins varying in abundance were identified, with Gly m 5 and Gly m 6 being the predominantly abundant allergens.

1. Introduction

Considering a desire for a healthy lifestyle, environmental awareness, and as an aversion to animal cruelty, consumers in recent years have been opting more toward a plant-based diet. Among others, plant-based milk alternatives have become one of the fastest-growing and emerging segments in the food industry (Sethi et al., 2016). In the recent years, there has been a notable increase and rapid spread in the consumption of plant-based milk mainly due to its nutritional value and numerous positive health benefits on the human body (Zujko and Witkowska, 2014). With 70–100% of people in east Asia suffering from lactose intolerance and around 65% of the world's population with a reduced ability to digest lactose, the reliance of plant-based milk has become more than a lifestyle choice, but an ideal alternative option to meet daily nutritional demands of the consumers with lactose intolerance and cow milk allergy (Mäkinen et al., 2016).

Among other plant sources including rice, oat, hemp, coconut, and

nuts, soy milk is one of the most widely used and common plant-based milk alternatives, especially among consumers with lactose intolerance and cow milk allergy (Sethi et al., 2016, Vanga and Raghavan, 2018). Soy milk, a colloidal solution extracted from ground soybeans in water, has its history dating back to more than 2000 years ago among the Asian civilizations. It is approved by the FDA as a health food and retains almost all of components of soy seeds (proteins, lipids and saccharides); moreover, it has no lactose and cholesterol and the protein composition is similar to that of cow milk. As a widespread alternative to cow milk, consumption of soy milk has been positively linked to numerous health benefits including reduced risks to cancer, cardiovascular and neurodegenerative disorders (Badger et al., 2005, Pistollato et al., 2018, Eslami and Shidfar, 2019). Besides high-quality proteins, soy milk also comprises a variety of nutrients, polyunsaturated fatty acids, saponins as well as phytosterols, soy lecithins, and isoflavones, (Ramdath et al., 2017) making it a functional beverage that has a multifold positive impact on human health. However, plant-based milk

* Corresponding authors at: Department of Biomedical Sciences, University of Padova, via U. Bassi 58/B, 35131 Padova, Italy (G. Arrigoni) and Department of Agronomy, Food, Natural Resources, Animals, and Environment, University of Padova, Via dell'Università 16, 35020 Legnaro, Italy (A. Masi).

E-mail addresses: ilaria.battisti@studenti.unipd.it (I. Battisti), leonardbarnabas.ebinezer@unipd.it (L.B. Ebinezer), giovanna.lomolino@unipd.it (G. Lomolino), antonio.masi@unipd.it (A. Masi), giorgio.arrigoni@unipd.it (G. Arrigoni).

¹ These authors equally contributed to this work.

substitute products also have negative health effects including low bioavailability of essential amino acids, minerals and vitamins, and oral health problems (Aydar et al., 2020). Especially proteins in soy milk could be allergenic, as around 14% of the consumers with cow milk allergy also have reactions against soy milk (Zeiger et al., 1999).

Over the past decade, proteomics has become a valuable tool to study the protein composition of foods (Piras et al., 2016) and it has been recently gaining popularity as a promising approach for the identification and quantification of allergenic proteins (Mainente et al., 2017, Marzano et al., 2020) as well as an innovative strategy in the field of food traceability and authenticity. Most of past proteomic research on soy and soy-related products aimed at characterizing soy proteins in seeds, with particular reference to a subset of proteins with allergenic potential, and establishing proteomics pipeline to study soybean seed filling and seed proteome (Hajduch et al., 2005, Gomes et al., 2014). Specifically, regarding the soy milk proteome, Capriotti et al. (2014) compared the protein profile of mature soybean seeds with different preparations of soybean milk using a shotgun proteomics approach. Nevertheless, their analysis is essentially qualitative and is limited to the description of proteins identified in seeds and soy milk, highlighting the similarity between the different samples. More recently, Al-Saedi et al., (2020) analyzed the protein profile of soy milk prepared by two different technological processes using two-dimensional polyacrylamide gel electrophoresis (2-DE). However, to the best of our knowledge, a comparison between the protein profiles of commercially available soy milks using a quantitative proteomic approach has not yet been carried out. The technological and thermal treatments to which the soy milk is subjected, both in the extraction phase and to extend the shelf-life, could influence the final composition of the proteins present in the commercial product. In fact, denaturation, coagulation and precipitation are phenomena that occur with different extent on soy milk proteins, depending on the intensity of the technological stabilization treatment of the product. These phenomena could affect not only the protein composition of the various products on the market but also the more or less accentuated presence of allergenic proteins. In the present study, we applied a label-free quantitative proteomic approach to 15 different commercially available soy milks with the aim to highlight possible variations in their protein composition and to estimate the amount of potential allergenic proteins present in the commercial products. The label-free proteomic approach used in this study relies on the relative quantification of proteins based on the precursor signal intensity generated by the peptides obtained upon proteolytic cleavage. The high-resolution, high mass accuracy mass spectrometer used to acquire the data, enables the extraction of peptide signals at the MS level, while MS/MS spectra and database search allow for the identification of the peptide sequences. Dedicated software packages are then used to align/normalize the chromatographic traces and extract the quantitative information from the different samples. The collected data were used to perform an unsupervised clustering of the analyzed commercial soy milks based on the most abundant proteins. Distinct clustering of soy milk samples observed in our study might reflect the different thermal treatments to which soy milks were subjected during manufacturing as well as encompass the variability due to the genetic background of the soy cultivars used for the production process.

2. Materials and methods

2.1. Protein precipitation from soy milk samples

All chemicals and reagents used in this study were of high purity grade purchased from Sigma-Aldrich, and Fluka, unless specified. Commercially available soy-based milk beverages used for this study were manufactured and commercialized by different companies and purchased at local supermarkets. Starting from 100 μ l of soy milk for each sample, proteins were precipitated using one volume of 72% trichloroacetic acid (TCA) added to four volumes of the soy milk samples.

Following incubation on ice for 10 min, samples were centrifuged at 15000 rpm for 15 min. After washing twice with two volumes of cold acetone (-20°C), the protein pellets were air-dried and suspended in buffer A (8 M urea, 100 mM Tris-HCl, pH 8.5). Protein concentration was quantified by Bradford assay as routinely carried out, using bovine serum albumin (BSA) as a standard.

2.2. Protein digestion

Protein digestion was carried out using the Filter-Aided Sample Preparation (FASP) method. Briefly, 40 μ g of proteins were mixed with Buffer A to a final volume of 200 μ l, loaded into centrifugal ultrafiltration units with a nominal molecular weight cutoff of 10 kDa (Vivacon 500, Sartorius) and centrifuged at 14000 rpm for 15 min. The flow-through was discarded and the filter was washed twice with 200 μ l of buffer A by centrifuging at 14000 rpm for 15 min. To reduce protein disulfide bridges, 200 μ l of 50 mM dithiothreitol (DTT) in buffer A was added and samples were incubated at 55°C for 30 min. After centrifugation, 100 μ l of 50 mM iodoacetamide (IAA) in buffer A was added to alkylate free thiol groups, and samples were incubated for 20 min in the dark. Filter units were washed twice with 100 μ l of buffer A, and then other two times with 100 μ l of 100 mM ammonium bicarbonate and 50 mM ammonium bicarbonate respectively, by centrifuging at 14000 rpm for 10 min. Proteins were digested at 37°C overnight with Sequencing Grade Modified Trypsin (Promega) in 50 mM ammonium bicarbonate, using an enzyme to protein ratio of 1:50. Peptides were collected by centrifugation at 14000 rpm for 10 min, followed by two washes with 100 μ l of 50 mM ammonium bicarbonate. Peptides samples were acidified with 1 μ l of formic acid (FA), and diluted to 1 ml with 0.1% FA, before desalting with C18 cartridges (Sep-Pak, C18, Waters) according to manufacturer's instructions. Finally, samples were dried under vacuum and stored at -20°C until LC-MS/MS analysis.

2.3. Liquid chromatography-tandem mass spectrometry analysis

Mass spectrometry analysis was performed with an LTQ Orbitrap XL (Thermo Fisher Scientific) interfaced with a nano-HPLC Ultimate 3000 (Dionex – Thermo Fisher Scientific). Samples were suspended in 80 μ l of 3% acetonitrile (ACN)/0.1% FA and for each sample 1 μ l (corresponding to 0.5 μ g of digested proteins) was loaded into a trap column (NanoEase Symmetry300, C18, 5 μ m, Waters) at a flow rate of 8 μ l/min. The peptide mixture was separated in a pico-frit capillary column (11 cm, 75 μ m internal ID, 15 μ m tip, New Objective) packed in-house with C18 material (Aeris Peptide 3.6 μ m XB-C18, Phenomenex), using a flow rate of 250 nl/min and a linear gradient of ACN/0.1% FA from 3 to 40% in 40 min. Ion source capillary was set to 200°C and spray voltage was fixed at 1.5–1.6 kV. The instrument operated in a data-dependent mode, with a full scan in the Orbitrap (60000 resolution) in the 300–1700 m/z range, followed by MS/MS scans of the 10 most intense ions acquired at lower resolution in the linear trap. A single LC-MS/MS analysis was performed for each of the 15 samples.

2.4. Database search and quantification

Raw MS data files were analyzed with MaxQuant (Cox and Mann, 2008) software v.1.5.1.2. Proteins search was performed against the Soybean section of the UniProt database (*Glycine max*, version 2018-06-21, 75,675 sequences), concatenated with the Cocoa section (*Theobroma cacao*, version 2018-06-21, 40,615 sequences), and the Almond section (*Prunus dulcis*, version 2018-06-21, 449 sequences). Trypsin was set as enzyme with up to 2 missed cleavages allowed, while carbamidomethylation of cysteine residues and oxidation of methionine were set as fixed and variable modifications, respectively. A false discovery rate (FDR) of 0.01, both at the peptide and protein level, was used to filter the results. A minimum of two peptides was required for protein identification. To evaluate the relative protein abundance within samples, we

used the Intensity Based Absolute Quantification (iBAQ) parameter calculated by the software. The iBAQ value associated with a protein correlates with the molar amount of that protein (Schwanhäusser et al., 2011) and therefore the estimated mass of a protein in each sample can be calculated from the corresponding iBAQ value multiplied by the theoretical molecular weight. This mass value in arbitrary units was then divided by the sum of all mass values in the sample to obtain the % mass for each of the identified proteins (Finka et al., 2015). To assess the presence of unsuspected post-translational modifications (PTMs) we analyzed our data also with the Error Tolerant Search feature of Mascot search engine (Matrix Science). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (www.proteomexchange.org) via the PRIDE partner repository with the dataset identifier PXD023052.

2.5. Statistical analysis

Statistical analysis was performed on % mass values (calculated as described above) with Past version 4.01. For principal component analysis, variance-covariance matrix was selected and empty cells were filled with zero. Row-wise bootstrapping was not applied. For cluster analysis, non-hierarchical K-means clustering algorithm was chosen, setting $K = 3$. For box plots, % mass values were used and the median was included. Values that were greater than $Q3 + IQR \times 1.5$ or less than $Q1 - IQR \times 1.5$ were marked as outliers and represented with dots (IQR : InterQuartile Range; $IQR = Q3 - Q1$).

2.6. Bioinformatic analyses

BLASTP tool (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) was used to recover the descriptions of the uncharacterized proteins. Initially, BLASTP was performed against UniProtKB/Swiss-Prot limited to *Glycine max* (taxid: 3847), *Prunus dulcis* (taxid: 3755), *Prunus dulcis* × *Prunus persica* (taxid: 472390), *Theobroma cacao* (taxid: 3641), *Glycine* (taxid: 3846), *Prunus* (taxid: 3754), *Theobroma* (taxid: 3640). To improve the recovery, we extended the analysis against *Fabaceae* (taxid: 3803), *Malvaceae* (taxid: 3629), *Rosaceae* (taxid: 3745), and finally to *Viridiplantae* clade. Sequence coverage and sequence identity cutoffs were set to 60% and 50%, respectively. Heat maps were generated using the % mass data with Heatmapper (<http://www.heatmapper.ca>), values were not scaled and hierarchical clustering was not computed. Gene ontology (GO) enrichment analysis was carried out with g:Profiler (<https://biit.cs.ut.ee/gprofiler/gost>) with Benjamini-Hochberg false discovery rate (FDR) significance threshold set at 0.05. GO enrichment was analyzed in EnrichmentMap and visualized in Cytoscape. Molecular function-based categorization of the identified proteins was carried out manually with the information available in UniProt (<https://www.uniprot.org/>). Amino acid content contributed by the 23 most abundant proteins in all the soy milk samples was deduced using the % mass data (Table S4) and the amino acid composition of each protein obtained using ExPASy ProtParam tool (<https://web.expasy.org/protparam/>). The estimated content of each amino acid in grams per 100 g of soy milk proteins was calculated as follows: (% mass of protein × % of amino acid)/100; this calculation was performed for all 23 proteins and for all 20 amino acids. The grand average of hydrophobicity (GRAVY) index was calculated with the GRAVY calculator tool (<http://www.gravycalculator.de>) using the amino acid sequences of the identified proteins. Classification of proteins as potential soy allergens is based on the data available in three databases, namely Allergome (<http://www.allergome.org>), COMPARE (<https://comparedatabase.org>), and AllerBase (<http://bioinfo.unipune.ac.in/AllerBase/Home.html>), selecting “*Glycine max*” as organism source.

3. Results and discussion

The production and consumption of soybean-related products have

increased significantly over the years due to its nutritional values, reported health benefits (Badger et al., 2005; Aydar et al., 2020), and soybean milk being an alternative to cow milk. With a multitude of commercial soy milks available, their protein composition after passing through rigorous manufacturing processes and commercialization is not well defined. In the present study, we used a quantitative proteomic approach to compare the soy milk proteome of 15 different commercially available soy milks. The drinks were bought at local supermarkets and randomly selected among the brands on sale, to cover most of the available commercialized types of products. The ingredients reported in the label of each drink are listed in Table S1: it can be noted that most of the soy milks used in this study contain only soy as source of proteins, while in few cases (samples 2, 3, 5, 7, 13 and 15) other potential sources of proteins were used during manufacturing. To keep into account the possible (although supposedly minor) contribution of these other protein sources, a dedicated database was compiled by concatenating the following databases retrieved in UniProt (www.uniprot.org): *Glycine max*, *Theobroma cacao*, and *Prunus dulcis*. We also checked for the presence of proteins possibly derived from other plants listed among the ingredients (such as barley, rye, chicory, apple and vanilla) but we could not identify any, suggesting that these ingredients were added mainly to provide a flavor or as a source of fibers. Accordingly, we then did not include these species in our database to limit the number of entries and reduce the risk for false positive identifications.

3.1. Composition of the soy milk proteomes

Using a label-free proteomics approach, 186 proteins in total were reliably identified and quantified (see Table S2, which contains all information to assess the reliability of protein and peptide identifications). Of the total identified, 170 proteins (91%) were from *Glycine max*, 9 from *Prunus dulcis* (5%), and 7 from *Theobroma cacao* (4%). To note that almond and cocoa were listed as ingredients only in sample 5 and sample 3, respectively. Among the 170 soybean proteins, there were 51 “uncharacterized”, however, performing a BLAST-based search we were able to recover the descriptions of 33 proteins. Only 18 soybean proteins remained therefore uncharacterized, which is around 10% of total identified proteins (Table S3). Using the Intensity Based Absolute Quantification (iBAQ) value and the theoretical molecular weight of each protein, we calculated the % mass of all identified proteins, as detailed in the Materials and Methods section and reported in Table S4. Taking into account the only two samples that contain other sources of proteins besides soy, we found that proteins from *P. dulcis* contributed to almost 10% to the total amount retrieved in sample 5, while those from *T. cacao* contributed to only about 0.3% of proteins present in sample 3. We cannot exclude that the contribution of *P. dulcis* proteins in sample 5 could be even more important, considering that a database with a very limited number of entries was available for this species at the time of the analysis. Our findings suggest that while almonds are added to the drink in appreciable amount and significantly contribute to the nutritional value of the product, cocoa (and therefore the proteins sourced from it) is likely added only to enhance/mask the beany flavor often associated with soy milk.

3.2. Physico-chemical properties and classification of the identified proteins

The theoretical molecular weight of the identified proteins ranges from 8 to 103 kDa, with 54% within 10 to 30 kDa, followed by 30% and 12% of the proteins in the range of 30–60 kDa and 60–100 kDa, respectively. Proteins less than 10 kDa accounted for around 4% (Table S5). For about 81% of the identified proteins the isoelectric point (pI) is within 4–7, while the rest have a pI value within 7–11 (Table S5). Around 85% of the proteins have a negative grand average of hydrophathy (GRAVY) value, indicating the prevalence of hydrophilic proteins in the analyzed soy milk samples (Table S5). The physico-chemical

properties of the soy milk proteins could largely depend on the technological processes used in manufacturing these drinks, especially heat, high hydrostatic pressure, pulsed electric field, ultra-high pressure homogenization, ultrafiltration among others. These processes have been known to denature, increase insolubility, dispersion, and viscosity of soy proteins (Giri and Mangaraj, 2012). Besides, ultrafiltration (with 20 kDa cutoff) of soy milk intended to remove low-molecular-weight anti-nutritional factors, including oligosaccharides, raffinose and stachyose, and phytic acid, could also result in the loss of low-molecular-weight proteins, which could partially explain the lower representation of proteins below 10 kDa observed in our study.

Manual categorization of all the identified proteins based on the molecular function information available in the UniProt database indicates that the majority of the proteins are related to binding, followed by nutrient reservoir activity, and oxidoreductase activity (Fig. 1A). Further, superimposing the % mass data (Table S6) onto the functional categorization, we observe that the highest contribution in terms of % mass is given by the proteins related to the category of nutrient reservoir activity, followed by endopeptidase inhibitor activity, lipid binding, and seed maturation (Fig. 1A). These data in turn reflect the protein composition of soybean which are known to be particularly rich in proteins with nutrient reservoir activity as around 70–83% of the soybean proteins are reported to have storage functions (Derbyshire et al., 1976). More recently, based on a proteomic analysis Yang et al. (2016) reported that over 81% of the total abundance in soybean seeds was represented by proteins with nutrient reservoir activity. This is also reported in previous proteomics studies on soy milk (Capriotti et al., 2014, Al-Saedi et al., 2020) and is consistent with our observation that nutrient reservoir activity-related proteins are highly abundant in soy milk.

Gene ontology (GO) enrichment analysis reveals that GO terms associated with oxidoreductase activity are significantly enriched (Fig. 1B), supporting the observation we made with molecular function-based categorization of identified proteins. Other GO terms are significantly enriched, namely enzyme inhibitor activity, binding, structural constituent of ribosome, protein folding, and translation. Apart from the storage proteins, the soybean contains around 20–30% of bioactive proteins that include beta-amylase, cytochrome c, lectin, lipoxygenase, urease, the Kunitz inhibitor of trypsin, and the Bowman-Birk inhibitor of chymotrypsin and trypsin (Friedman and Brandon, 2001). As many of the soy proteins are multi-functional they could be categorized differently. However, several protein categories (notably, nutrient reservoir activity, binding, translation, and enzyme regulator activity, and anti-oxidant activity) found in the previous proteomics study on soy milk (Capriotti et al., 2014) were also significantly enriched in our study.

3.3. Comparison of the different soy milk proteomes

To compare the different soy milk samples and possibly highlight similarities and differences in terms of protein composition, we carried out a principal component analysis (PCA) using the % mass values of all identified proteins. PC1 and PC2 explained almost 70% of total variability and three distinct clusters could be observed (Fig. 2). We also performed a non-hierarchical cluster analysis with the K-means algorithm and observed identical clustering of the soy milk samples. Group A (red) includes soy milk samples 3, 6, 8, 9, 11, 12, 14, and 15, group B (light blue) comprises of milks 1, 2, 5, 13, while samples 4, 7, and 10 cluster in group C (green).

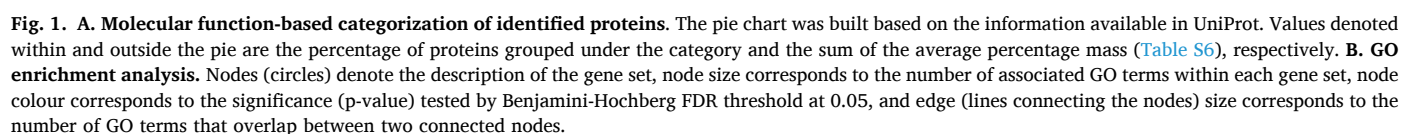
To interpret PCA and clustering results, we considered the first three principal components (PCs), which explain 80% of the variance (Table S7). Plotting each PC against each other, we noticed that PC1 and PC2 discern the three groups better than PC3. To figure out which proteins mainly determine the PCs, we considered only the proteins with at least two non-zero loadings (to one decimal point) (Table S7). PC1 has a negative contribution from beta-conglycinins and a positive contribution from glycinins and cocoa proteins; prunins give a null contribution. PC2 is negatively determined by most of seed storage proteins like

glycinins and beta-conglycinins, while prunins and oleosins contribute positively. Finally, almond proteins and soy seed storage proteins negatively contribute to PC3, although oleosins and cocoa proteins provide a positive contribution. It is credible that different soy milk groups could reflect the different production processes and/or the soy cultivars used for milk preparation. Interestingly, soy milks containing cocoa (sample 3) and almonds (sample 5) belong to different groups. Sample 3 clusters in group A, together with most of the other milks, supporting the hypothesis that cocoa is added principally for taste reasons rather than for its nutritional value, affirming the observation we made with % mass data. Conversely, sample 5 is part of group B but stands out from all the other samples, thus reflecting the remarkable contribution of almond on the total protein content.

Further, based on the average of % mass values of the identified proteins across the 15 different soy milk samples, we observed that the 23 most abundant soybean proteins contribute to 90% of the total mass (Table S8). Considering that we identified in total 186 proteins, it means that 87.6% of these are present in small amounts or even just in traces. Among these 23 proteins, there were multiple isoforms of glycinins, beta-conglycinins, oleosins, sucrose binding proteins, and late embryogenesis abundant proteins, with a beta-conglycinin (P0D016) having the highest average % mass. As also previously observed by Capriotti et al. (2014), glycinins and beta-conglycinins are the most abundant proteins in the soy milk samples analyzed in our study. Based on the sedimentation coefficients, soybean protein consists of four fractions (2S, 7S, 11S, and 15S), of which the 7S (beta-conglycinin) and 11S (glycinin) fractions account for about 70% of the total seed protein (Fukushima, 2001). To easily visualize differences between samples, we computed a box plot with average % mass values (Fig. 3A).

The most variable proteins in amount are beta-conglycinins, while glycinins and oleosins have less variability between different samples. The remaining proteins are less abundant compared to the others, as a consequence, their variability is rather limited. A total of 13 outliers are present in 7 of the 23 proteins: it is noteworthy that 7 of them are distributed in three glycinins. Regarding the milk samples, the highest number of outliers belongs to sample 5: this particular soy milk stands out also in PCA analysis (Fig. 2) and this is probably due to the presence of almond, which impacts on the % mass of soy proteins computed for this sample. Notably, samples 4, 7, and 10 show outlier values for the three glycinins, and this could be one of the reasons why they constitute one specific group of samples (group C) in the PCA plot. As a confirmation, we generated a heat map with % mass values of the 23 most abundant proteins (Fig. 3B) and of all the 186 identified proteins (Fig. S1), organized according to the clusters observed in the PCA plots. Some differences are clearly visible across groups, especially for proteins that are characterized by a high % mass value (that are substantially glycinins and beta-conglycinins). In particular, beta-conglycinins F7J077 and P25974 are present in a larger quantity in samples belonging to group A, while glycinin P02858 is higher in soy milk of group C. It is worth to note that the amount of beta-conglycinins F7J077 and P25974 is very similar within each sample (Fig. 3B and Table S8): these proteins are two different subunits of the same protein complex, therefore their presence in similar amount supports the reliability of the quantitative data we have generated.

The heat treatments, to which soy milk is subjected, have the purpose of prolonging the shelf-life, reducing the microbial load and inactivating the trypsin inhibitor proteins which, in their native form, constitutes an anti-nutritional factor. However, heat treatment can compromise the colloidal stability of some proteins by causing coagulation and precipitation (Manassero et al., 2016). The phenomenon of protein instability may be due to the complex quaternary structure of the 7S and 11S globular proteins of soy milk (Fukushima, 2001). Fractions 7S and 11S are made up of subunits that undergo reversible associations and dissociations. Heat causes dissociation of the 7S and 11S subunits and interaction between the dissociated subunits that form aggregates and precipitates. The extent of the formation of precipitates depends on the



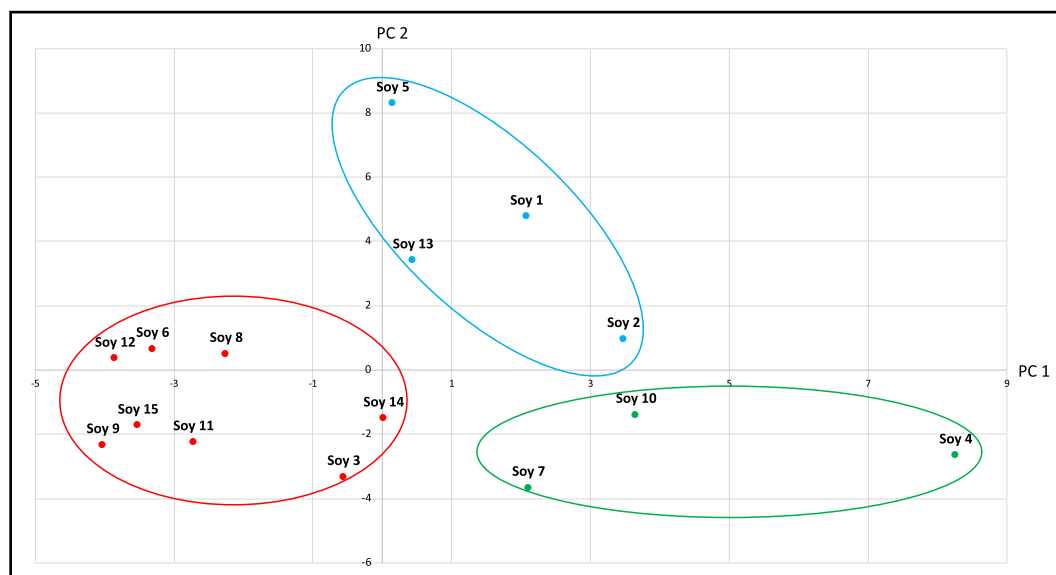


Fig. 2. PCA plot of soy milk samples. Colouring of the soy milk samples is based on the cluster analysis result.

time/temperature binomial relationship of the heat treatment. In fact, the optimal thermal conditions guarantee the unfolding and dissociation of soy milk proteins, exposing more ϵ -amino groups of lysine and in general making soy milk proteins more digestible. Conversely, more drastic treatments cause cross-linking of unfolded peptide chains, precipitation, and decreased nutritional value. In the light of these notions, our observations and the unsupervised clustering of samples in three distinct groups might reflect the different thermal treatments to which soy milks were subjected during manufacturing.

To assess the presence of unsuspected post-translational modifications (PTMs), raw data files were analyzed with the Error Tolerant Search feature of Mascot search engine by selecting the most abundant proteins identified in each sample. The outcome of this analysis, which exploits a smaller size database to perform a search by taking into account a comprehensive list of PTMs, highlighted the presence of deamidation of asparagine/glutamine residues and carbamylation of lysine residues as the most abundant PTMs in our samples. A new search done with MaxQuant adding deamidation (N/Q) and carbamylation (K) as variable modifications, showed that the new data did not have a significant impact on our downstream analysis.

While the deamidation of asparagine/glutamine residues is a known physiological process, the rate of conversion between Asn/Gln and Asp/Glu is influenced by several factors among which are pH values and temperatures. On the other hand, carbamylation of lysine residues is a very well-known irreversible modification that is usually observed when proteins are treated at temperatures greater than 37 °C in the presence of urea. In the protocol for sample preparation applied in this study urea is used at high concentration (8 M) but at room temperature. It is conceivable that such modification could be induced by the industrial processes to which the soy milks were subjected, however a dedicated study should be performed to further investigate this intriguing aspect. Since such PTMs may be the result of technological (mainly thermal) processes related with soy milk production and stabilization, their occurrence and abundance across the analyzed samples, are *per se* indicative of alterations, and possible loss of nutritional value. Therefore, our findings also highlight the importance and novelty of applying the proteomic pipeline as a tool to assess the nutritional value of artificial milks (and generally of processed foodstuff) in terms of the degree of deviations from the raw protein characteristics.

3.4. Nutritional value in terms of computed amino acid content

Based on the % of amino acid and the % mass, we determined the grams of amino acids contributed by the 23 most abundant proteins (per 100 g of soy milk proteins) (Fig. 4; Table S9). All soy milk samples have a relatively lower content of methionine and tryptophan among the essential amino acids (Table S9). Soy milk samples within each group, as defined by the PCA, show different amounts of essential amino acids (EAA) and non-essential amino acids (NEAA) (Fig. 4A). In particular, samples belonging to group A display some differences among them, but overall they have an amino acid content higher than the average. The amount of EAA of soy milks of group B is close to the average for samples 1, 2 and 13 and much lower for sample 5, a similar trend is also observable for NEAA. This data does not come as a surprise, considering that the 23 most abundant proteins used for this elaboration are all from soybean, while sample 5 contains almost 10% of proteins from *P. dulcis*. It is conceivable that almond proteins could provide a considerable contribution to the amino acid content that may influence the nutritional properties of the drink. Indeed, if we add to sample 5 the contribution of the three most abundant *P. dulcis* proteins (Fig. 4A, grey bar), the amount of both EAA and NEAA increases considerably, giving final values that are more in line with those found for the milks belonging to the same group. On the contrary sample 3, wherein cocoa has been supplemented, has an amino acid profile in line with the average, again supporting the hypothesis that cocoa is added principally for taste reasons. Finally, group C includes soy milks with highly similar amino acid content, especially for NEAA. Also to be noted is that these samples have EAA and NEAA content respectively lower and higher compared to the average. Conversely, samples of group A and group B exhibit the same trend between EAA and NEAA. Hence, PCA clusters go together with different amino acid composition, once again probably reflecting the different commercial-scale chemical/processing treatments used during manufacturing. Several of these (extensively reviewed by Giri and Mangaraj, 2012) are targeted to inactivate anti-nutrients compounds and their associated “beany” off-flavor. However, as mentioned above, these processing methods also influence protein stability, often leading to protein denaturation/precipitation and decline in specific amino acid content (Giri and Mangaraj, 2012). On the other hand, we cannot rule out the possibility that the dissimilarities observed in amino acid content between the analyzed soy milk samples could be attributed to the intrinsic differences in the soybean genotype used for the preparation. Large differences in protein profile and amino

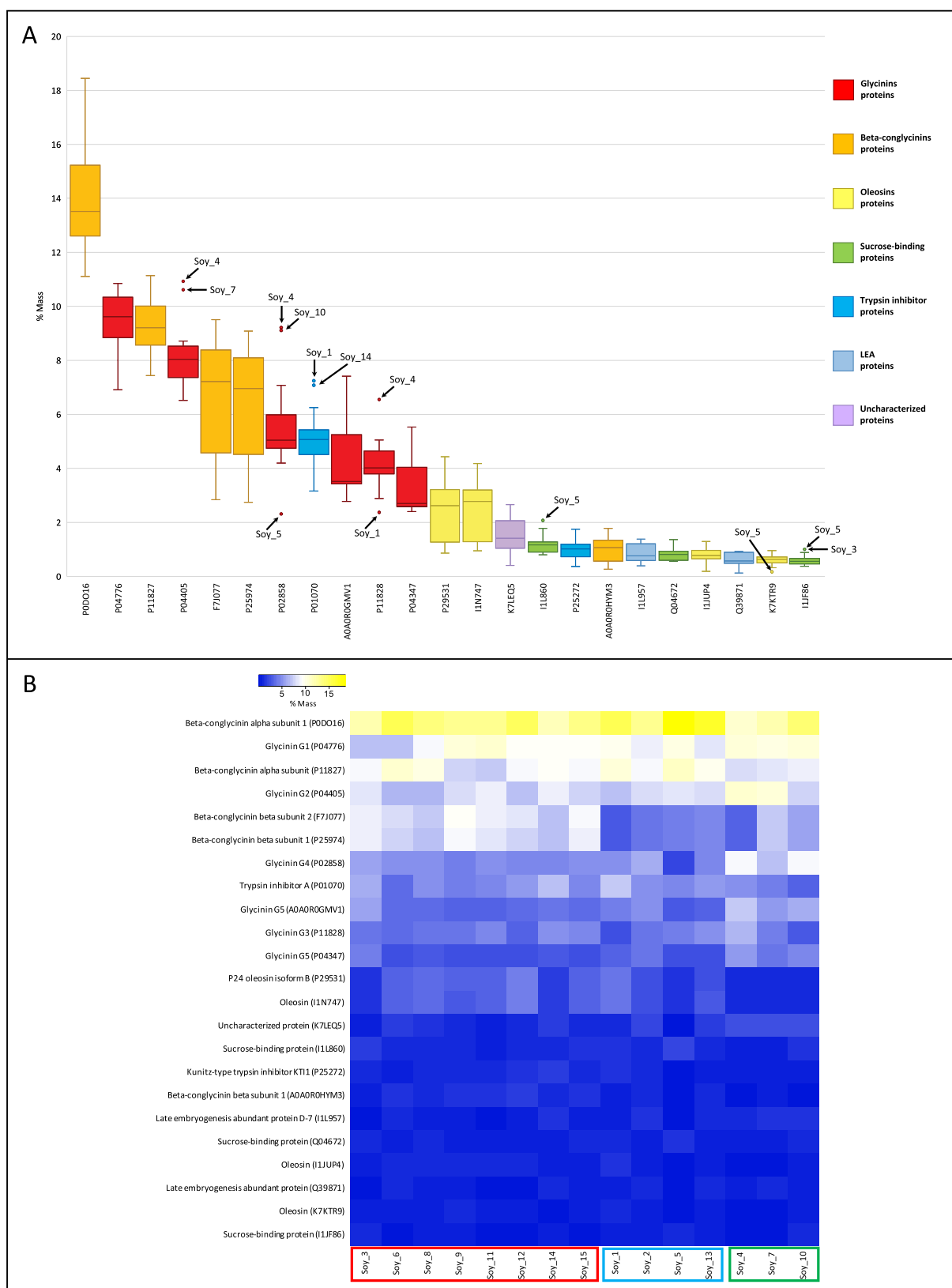


Fig. 3. A. Box plot showing the distribution of the percentage mass values of the 23 most abundant proteins. Proteins are identified by their UniProt ID and coloured according to the protein category. Outliers are displayed with dots and highlighted with arrows. **B. Heat map of the 23 most abundant proteins.** Soy milk samples are arranged based on the clusters observed in PCA, values denote % mass and are not scaled.

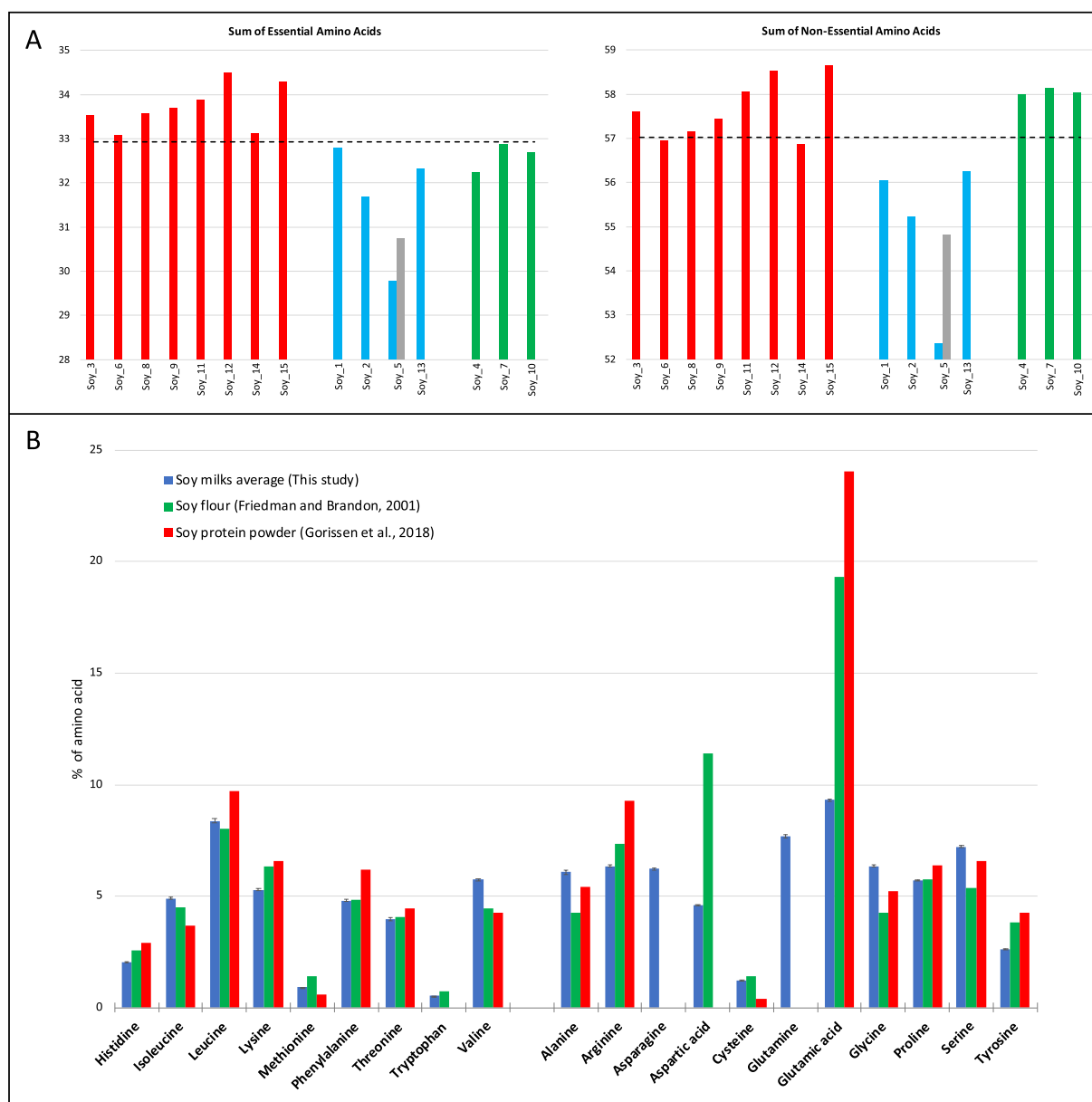


Fig. 4. A. Essential amino acid (EAA) and non-essential amino acid (NEAA) contributed by the 23 most abundant proteins. The sum of the essential and non-essential amino acids was calculated from the amino acid composition and percentage mass (Table S9) and the black dotted line denotes the average. Grey bars indicate the sum of EAA and NEAA considering also the three most abundant proteins from *P. dulcis* in sample 5. Colouring of the soy milk samples is based on the clustering observed in PCA. **B. Comparison of the average amino acid content in the soy milk samples with the reported literature.** Values from the publicly available supplementary information associated with the indicated publications were used for the comparison with the amino acid content computed from amino acid composition and percentage mass data. All values are expressed in percentage (Table S9). Error bars represent $\pm$ standard error. For all samples only soy proteins were taken into consideration.

acid content, together with other seed composition parameters, have been reported among soybean genotypes. These differences are influenced by climatic factors (Assefa et al., 2018) and were shown to be governed by a large number of quantitative trait loci (QTLs) revealed by several genome-wide association studies (GWAS) (Zhang et al., 2018).

Further, to assess whether the amino acid composition we estimated starting from the proteomics data is reliable, we compared the average of the amino acid content in the soy milk samples with the amino acid content reported in protein samples sourced from soybean (Fig. 4B). In particular, we considered amino acid composition of proteins derived from soy flour (Friedman and Brandon, 2001) and from a commercially available soy protein powder (Gorissen et al., 2018). In order to make the data comparable, we expressed amino acid content in terms of percentage values (Table S9). The estimated amino acid content of the

analyzed soy milk samples is comparable with values reported in literature, especially with those published by Friedman and Brandon (2001). Some minor differences are evident between our study and published data, which can be probably explained by the different technological processes the soybeans underwent for milk, flour or powder production, as well as by the method used for amino acid content determination. Indeed, it is interesting to note that major differences can be observed only for the content of aspartic acid, glutamic acid, asparagine and glutamine, with the latter two amino acids not reported in the studies we have used as comparison. We used a computational method to assess the amino acid composition, based on the sequence of the identified soy proteins, while the amino acid profile in the above-mentioned studies was experimentally determined. The most common experimental procedure used to evaluate the amino acid profile is through acidic

hydrolysis: this process implies the decomposition of tryptophan and the conversion of asparagine and glutamine into aspartic acid and glutamic acid, respectively. As a consequence, it is not possible to estimate the amount of these amino acids, unless other techniques are adopted. Our method is not affected by this limitation; thus we could estimate the contribution of all the 20 amino acids. As proof of this, if we sum the estimated amount of asparagine and aspartic acid, their value is again in line to the amount of aspartic acid measured by [Friedman and Brandon \(2001\)](#). The same applies for glutamine and glutamic acid, thus indicating once again that the approach used in our study to calculate the % mass through the iBAQ parameter is reliable and the method we used to compute the amino acid content in soy milk samples is reliable to the experimentally obtained data.

3.5. Allergens identified in the analyzed soy milk samples

While soybean milk is known for its high nutritional value, it is also considered as one of the most allergenic foods and is listed among the “big twelve” allergenic foods in Europe (European Food Safety Authority (EFSA)). Allergic response to soybean products is attributed to at least one of the 44 known allergenic proteins that are IgE-binding. The major allergens of soy bean are seed storage proteins 11S glycinin (Gly m 6), 7S beta-conglycinin (Gly m 5), the oil body-associated Gly m Bd 30 K, Gly m Bd 28 K, and Gly m 3.6 ([Yang et al., 2011](#)). Allergens Gly m 5 and Gly m 6 are associated with severe allergic reactions ([Holzhauser et al., 2009](#)), whilst Gly m 2, Gly m 3, and Gly m 4 are highly cross-reactive. Based on the most recent version of the allergen database ([www.allergenonline.org](#)), there are 44 entries listed as potential allergens in soybean. Among these, 20 allergens were identified in this study ([Table 1](#)), which were predominantly glycinins (G1, G2, G3, G4, G5), beta-conglycinin subunits, 2S albumins, and trypsin inhibitors.

As updated in the official allergen database of the Allergen Nomenclature Subcommittee of the World Health Organization and International Union of Immunological Societies, there are eight verified

Table 1
List of allergens identified in the soy milk samples. UniProt ID, protein description, WHO/IUIS name and the average percentage mass calculated as described in Material and Methods section are listed.

UniProt ID	Description	WHO/IUIS Name	Average % mass
P0D016	Beta-conglycinin alpha subunit 1	Gly m 5	13.964
P04776	Glycinin G1	Gly m 6.0101	9.398
P11827	Beta-conglycinin alpha subunit	Gly m 5	9.240
P04405	Glycinin G2	Gly m 6.0201	8.124
F7J077	Beta-conglycinin beta subunit 2	Gly m 5	6.580
P25974	Beta-conglycinin beta subunit 1	Gly m 5.0301	6.405
P02858	Glycinin G4	Gly m 6	5.549
P01070	Trypsin inhibitor A	Gly m TI	5.101
P11828	Glycinin G3	Gly m 6.0301	4.179
P04347	Glycinin G5	Gly m 6.0501	3.301
P25272	Kunitz-type trypsin inhibitor KTI1	Gly m TI	0.968
Q39871	Late embryogenesis abundant protein	Gly m EAP	0.623
P19594	2S albumin	Gly m 8	0.558
P05046	Lectin	Gly m Agglutinin	0.501
Q39846	Seed biotin-containing protein SBP65	Gly m 7	0.296
P25273	Kunitz-type trypsin inhibitor KTI2	Gly m TI	0.117
C6T588	Bet v 1 domain-containing protein	Gly m 4	0.093
P24337	Hydrophobic seed protein	Gly m 1	0.054
Q9ZNZ4	Napin-type 2S albumin 1	Gly m 2S albumin	0.013
B1ACD5	Kunitz trypsin protease inhibitor	Gly m TI	0.003

soybean allergens (Gly m 1–8) ([WHO/IUIS Allergen Nomenclature Subcommittee, 2018](#)). In the soy milk samples analyzed, we did not identify allergens Gly m 2 and 3, while Gly m 4, one of the most prominent soy allergens that is also associated with birch pollen allergy was found in the range of 0.01–0.2% of total mass ([Table S10](#)), with an average value of 0.09%. Gly m 5 and Gly m 6 associated with severe allergic reactions have also been recently shown to be related to food-dependent exercise-induced anaphylaxis ([Hayashi et al., 2020](#)). Alarmingly, both Gly m 5 and Gly m 6 are among the most abundant proteins in the analyzed soy milk samples ([Table 1](#)). Considering the significance of soy allergens, several studies in recent years have been focussed on the development of diagnostic techniques and methods to identify and characterize allergenic proteins in soybean-based food products ([Cabanillas et al., 2018](#)). While there is no universal method for quantifying soybean allergens, mass spectrometry and related proteomic techniques have been relied upon predominantly to capitalize on sensitivity, accuracy, and reproducibility ([Yoshimitsu et al., 2019](#)). Whilst a highly sensitive multiple reaction monitoring (MRM)-based mass spectrometric method targeted to quantify peptides of the prominent allergens would be the ideal technique to investigate further into the soybean allergens, protein profiling like the present study can provide an overview of the extent of allergens present in commercial soy milk. [Capriotti et al. \(2014\)](#) identified 10 allergens in soy milk samples, of which, 8 (except the profilins) were also identified in our study. Albeit the manufacturing and processing methods are specifically targeted to degrade/inactivate the allergens and anti-nutrients like lipoxygenase and trypsin inhibitors ([Giri and Mangaraj, 2012](#)), they are still being detected in protein profiling studies ([Capriotti et al., 2014](#)), including commercialized soy milks as in the present study.

Whilst functional epitopes (amino acids accounting for most of the binding energy to IgE) of several predominant soy allergens are not yet known, [Kern et al. \(2019\)](#) have identified more than 400 potential epitope motifs in soybean proteins and showed that among the 83 peptides (representing the 42 most frequently found epitope candidates), majority of them were bound by antibodies, IgE, or by both. Except for Gly m 4 ([Husslik et al., 2016](#)), the conformational IgE epitopes of the predominantly found allergens in commercialized soy milks and soy products remain largely unknown. Apart from a molecular modeling-based simulation of the temperature and pressure-induced unfolding in soy allergen Gly m 4 ([Vanga et al., 2019](#)), there is very limited understanding on the molecular level effects induced by processing methods on the conformational epitopes of soy allergens. Hence, the allergens detected in the commercial soy milks in our study warrants further research, especially to verify if they are still immunogenic and what other technological improvements could be made to completely remove or inactivate these proteins without exacerbating the production cost. Notwithstanding the inadequate information on functional and conformational epitopes of allergens, remarkable progress has been made in the recent years with respect to developing technologies, techniques and processing methods to reduce the allergenicity of food products, including soy proteins ([Peñas et al., 2011](#)). There is also mounting evidence for reduced allergenicity in fermented soy milk and products ([Meinlschmidt et al., 2016](#)), which could be one of the potential processing methods to reduce allergenicity, making soy milk and soy products even safer for consumers.

4. Conclusions

The present study is the first of its kind to analyze the protein composition of commercial soy milk samples using a label-free quantitative proteomics approach. Results indicate that the protein composition and the consequent nutritional value, varies among the commercial soy milks. Samples could be categorized based on the relative abundance of proteins, probably reflecting the different technological/thermal treatments they undergo during manufacturing or the different soybean genotypes used for their production. The amino acid content

indirectly estimated from our proteomics data not only gives results that are comparable to values reported in literature, but it does not suffer of the typical limitation of acidic hydrolysis which does not allow to measure the amount of tryptophan, asparagine, and glutamine. All analyzed commercial soy milks show a relative lower content of histidine, methionine, tryptophan and cysteine, highlighting the opportunity to fortify commercial soy milks, in addition to the ongoing molecular breeding and genetic engineering efforts to improve the intrinsic nutritional content of the soybean. The immunogenicity of the allergens detected in the commercial soy milks in our study needs to be verified before deeming the otherwise very beneficial functional beverage completely safe for consumers.

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.

Acknowledgments

This study was funded by the University of Padova (Grant number: BIRD189887/18 to GA). LBE was supported by the postdoctoral grant by the University of Padova. The authors acknowledge the support provided by “Cassa di Risparmio di Padova e Rovigo” Holding (Cariparo) for funding the acquisition of the LTQ Orbitrap XL mass spectrometer.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.foodchem.2021.129299>.

References

- Al-Saedi, N., Agarwal, M., Ma, W., Islam, S., & Ren, Y. (2020). Study on effect of extraction techniques and seed coat on proteomic distribution and cheese production from soybean milk. *Molecules*, 25(14), 3237. <https://doi.org/10.3390/molecules25143237>.
- Assefa, Y., Bajjalieh, N., Archontoulis, S., Casteel, S., Davidson, D., Kovács, P., ... Ciampitti, I. A. (2018). Spatial Characterization of Soybean Yield and Quality (Amino Acids, Oil, and Protein) for United States. *Scientific Reports*, 8(1), 14653. <https://doi.org/10.1038/s41598-018-32895-0>.
- Aydar, E. F., Tutuncu, S., & Ozelcelik, B. (2020). Plant-based milk substitutes: Bioactive compounds, conventional and novel processes, bioavailability studies, and health effects. *Journal of Functional Foods*, 70, Article 103975. <https://doi.org/10.1016/j.jff.2020.103975>.
- Badger, T. M., Ronis, M. J. J., Simmen, R. C. M., & Simmen, F. A. (2005). Soy Protein Isolate and Protection Against Cancer. *Journal of the American College of Nutrition*, 24(2), 146S–149S. <https://doi.org/10.1080/07315724.2005.10719456>.
- Cabanillas, B., Jappe, U., & Novak, N. (2018). Allergy to Peanut, Soybean, and Other Legumes: Recent Advances in Allergen Characterization, Stability to Processing and IgE Cross-Reactivity. *Molecular Nutrition & Food Research*, 62(1). <https://doi.org/10.1002/mnfr.201700446>.
- Capriotti, A. L., Caruso, G., Cavaliere, C., Samperi, R., Stampachiachiere, S., Zenezini Chiozzi, R., & Laganà, A. (2014). Protein Profile of Mature Soybean Seeds and Prepared Soybean Milk. *Journal of Agricultural and Food Chemistry*, 62(40), 9893–9899. <https://doi.org/10.1021/jf5034152>.
- Cox, J., & Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology*, 26(12), 1367–1372. <https://doi.org/10.1038/nbt.1511>.
- Ramadh, D., Padhi, E., Sarfaraz, S., Renwick, S., & Duncan, A. (2017). Beyond the cholesterol-lowering effect of soy protein: A review of the effects of dietary soy and its constituents on risk factors for cardiovascular disease. *Nutrients*, 9(4), 324. <https://doi.org/10.3390/nu9040324>.
- Derbyshire, E., Wright, D. J., & Boulter, D. (1976). Legumin and vicilin, storage proteins of legume seeds. *Phytochemistry*, 15(1), 3–24.
- Eslami, O., & Shidfar, F. (2019). Soy milk: A functional beverage with hypocholesterolemic effects? A systematic review of randomized controlled trials. *Complementary Therapies in Medicine*, 42, 82–88. <https://doi.org/10.1016/j.ctim.2018.11.001>.
- Finka, A., Sood, V., Quadroni, M., Rios, P. D. L., & Goloubinoff, P. (2015). Quantitative proteomics of heat-treated human cells show an across-the-board mild depletion of housekeeping proteins to massively accumulate few HSPs. *Cell Stress & Chaperones*, 20(4), 605–620. <https://doi.org/10.1007/s12192-015-0583-2>.
- Friedman, M., & Brandon, D. L. (2001). Nutritional and health benefits of soy proteins. *Journal of Agricultural and Food Chemistry*, 49(3), 1069–1086. <https://doi.org/10.1021/jf0009246>.
- Fukushima, D. (2001). Recent Progress in Research and Technology on Soybeans. *Food Science and Technology Research*, 7(1), 8–16. <https://doi.org/10.3136/fstr.7.8>.
- Giri, S. K., & Mangaraj, S. (2012). Processing Influences on Composition and Quality Attributes of Soy milk and its Powder. *Food Engineering Reviews*, 4(3), 149–164. <https://doi.org/10.1007/s12393-012-9053-0>.
- Gomes, L. S., Senna, R., Sandim, V., Silva-Neto, M. A. C., Perales, J. E. A., Zingali, R. B., ... Fialho, E. (2014). Four conventional soybean [Glycine max (L.) Merrill] seeds exhibit different protein profiles as revealed by proteomic analysis. *Journal of Agricultural and Food Chemistry*, 62(6), 1283–1293. <https://doi.org/10.1021/jf404351g>.
- Gorissen, S. H. M., Crombag, J. J. R., Senden, J. M. G., Waterval, W. A. H., Bierau, J., Verdijk, L. B., & van Loon, L. J. C. (2018). Protein content and amino acid composition of commercially available plant-based protein isolates. *Amino Acids*, 50(12), 1685–1695. <https://doi.org/10.1007/s00726-018-2640-5>.
- Hajdich, M., Ganapathy, A., Stein, J. W., & Thelen, J. J. (2005). A Systematic Proteomic Study of Seed Filling in Soybean. Establishment of High-Resolution Two-Dimensional Reference Maps, Expression Profiles, and an Interactive Proteome Database. *Plant Physiology*, 137(4), 1397–1419. <https://doi.org/10.1104/pp.104.056614>.
- Hayashi, M., Pawankar, R., Yamanishi, S., & Itoh, Y. (2020). Food-dependent exercise-induced anaphylaxis to soybean: Gly m 5 and Gly m 6 as causative allergen components. *The World Allergy Organization Journal*, 13(7), 100439. <https://doi.org/10.1016/j.waojou.2020.100439>.
- 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>.
- Husslik, F., Nürnberg, J., Seutter von Loetzen, C., Mews, T., Ballmer-Weber, B. K., Kleine-Tebbe, J., ... Schiller, D. (2016). The conformational IgE epitope profile of soya bean allergen Gly m 4. *Clinical and Experimental Allergy: Journal of the British Society for Allergy and Clinical Immunology*, 46(11), 1484–1497. <https://doi.org/10.1111/cea.12796>.
- Kern, K., Havenith, H., Delaroque, N., Rautenberger, P., Lehmann, J., Fischer, M., ... Szardenings, M. (2019). The immunome of soy bean allergy: Comprehensive identification and characterization of epitopes. *Clinical and Experimental Allergy: Journal of the British Society for Allergy and Clinical Immunology*, 49(2), 239–251. <https://doi.org/10.1111/cea.13285>.
- Mainente, F., Simonato, B., Pasini, G., Franchin, C., Arrigoni, G., & Rizzi, C. (2017). Hen egg white lysozyme is a hidden allergen in Italian commercial ciders. *Food Additives & Contaminants: Part A*, 34(2), 145–151. <https://doi.org/10.1080/19440049.2016.1265673>.
- 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>.
- Manassero, C. A., Vaudagna, S. R., Sancho, A. M., Anón, M. C., & Speroni, F. (2016). Combined high hydrostatic pressure and thermal treatments fully inactivate trypsin inhibitors and lipoxigenase and improve protein solubility and physical stability of calcium-added soymilk. *Innovative Food Science & Emerging Technologies*, 35, 86–95. <https://doi.org/10.1016/j.ifset.2016.04.005>.
- Marzano, V., Tilocca, B., Fiocchi, A. G., Vernocchi, P., Levi Mortera, S., Urbani, A., ... Putignani, L. (2020). Perusal of food allergens analysis by mass spectrometry-based proteomics. *Journal of Proteomics*, 215, 103636. <https://doi.org/10.1016/j.jprot.2020.103636>.
- Meinlschmidt, P., Eisner, P., & Schweiggert-Weisz, U. (2016). Soy protein hydrolysates fermentation: Effect of debittering and degradation of major soy allergens. *LWT-Food Science and Technology*, 71, 202–212. <https://doi.org/10.1016/j.lwt.2016.03.026>.
- Peñas, E., Gomez, R., Frias, J., Baeza, M. L., & Vidal-Valverde, C. (2011). High hydrostatic pressure effects on immunoreactivity and nutritional quality of soybean products. *Food Chemistry*, 125(2), 423–429. <https://doi.org/10.1016/j.foodchem.2010.09.023>.
- Piras, C., Roncada, P., Rodrigues, P. M., Bonizzi, L., & Soggiu, A. (2016). Proteomics in food: Quality, safety, microbes, and allergens. *Proteomics*, 16(5), 799–815. <https://doi.org/10.1002/pmic.201500369>.
- Pistollato, F., Iglesias, R. C., Ruiz, R., Aparicio, S., Crespo, J., Lopez, L. D., ... Battino, M. (2018). Nutritional patterns associated with the maintenance of neurocognitive functions and the risk of dementia and Alzheimer's disease: A focus on human studies. *Pharmacological Research*, 131, 32–43.
- Schwanhäusser, B., Busse, D., Li, N. A., Dittmar, G., Schuchhardt, J., Wolf, J., ... Selbach, M. (2011). Global quantification of mammalian gene expression control. *Nature*, 473(7347), 337–342. <https://doi.org/10.1038/nature10098>.
- 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>.
- Vanga, S. K., & Raghavan, V. (2018). How well do plant based alternatives fare nutritionally compared to cow's milk? *Journal of Food Science and Technology*, 55(1), 10–20. <https://doi.org/10.1007/s13197-017-2915-y>.
- Vanga, S. K., Wang, J., Singh, A., & Raghavan, V. (2019). Simulations of Temperature and Pressure Unfolding in Soy Allergen Gly m 4 Using Molecular Modeling. *Journal of Agricultural and Food Chemistry*, 67(45), 12547–12557. <https://doi.org/10.1021/acs.jafc.9b05140>.

- WHO/IUIS Allergen Nomenclature Sub-Committee. (2018). Allergen nomenclature. Retrieved from <http://www.allergen.org/search.php>.
- Yang, A., Yu, X., Zheng, A., & James, A. T. (2016). Rebalance between 7S and 11S globulins in soybean seeds of differing protein content and 11SA4. *Food Chemistry*, 210, 148–155. <https://doi.org/10.1016/j.foodchem.2016.04.095>.
- Yang, W. W., De Meijia, E. G., Zheng, H., & Lee, Y. (2011). Soybean Allergens: Presence, Detection and Methods for Mitigation. *Soybean and Health*, 20, 433–464. <https://doi.org/10.5772/17858>.
- Yoshimitsu, M., Kiyota, K., Kajimura, K., & Yamano, T. (2019). Development of an LC-MS/MS-based analytical method for quantification of soybean allergen Gly m 4 in soybean grains and processed foods. *Food and Agricultural Immunology*, 30(1), 25–33. <https://doi.org/10.1080/09540105.2018.1540553>.
- Zeiger, R. S., Sampson, H. A., Bock, S. A., Burks, A. W., Harden, K., Noone, S., ... Wilson, G. (1999). Soy allergy in infants and children with IgE-associated cow's milk allergy. *The Journal of Pediatrics*, 134(5), 614–622.
- Zhang, J., Wang, X., Lu, Y., Bhusal, S. J., Song, Q., Cregan, P. B., ... Jiang, G.-L. (2018). Genome-wide Scan for Seed Composition Provides Insights into Soybean Quality Improvement and the Impacts of Domestication and Breeding. *Molecular Plant*, 11(3), 460–472.
- Zujko, M. E., & Witkowska, A. M. (2014). Antioxidant Potential and Polyphenol Content of Beverages, Chocolates, Nuts, and Seeds. *International Journal of Food Properties*, 17(1), 86–92. <https://doi.org/10.1080/10942912.2011.614984>.

