Unraveling the Impact of Tempeh Fermentation on Protein Nutrients: An In Vitro Proteomics and Peptidomics Approach

Jia Yee Wu[1,2], Sheena Wee[1], Siok Ghee Ler[1], Christiani Jeyakumar Henry[2], Jayantha Gunaratne[1,3*]

[1] Institute of Molecular and Cell Biology (IMCB), Agency for Science, Technology and Research (A*STAR), 61 Biopolis Drive, Singapore 138673, Singapore

[2] Clinical Nutrition Research Centre (CNRC), Singapore Institute of Food and Biotechnology Innovation (SIFBI), Agency for Science, Technology and Research (A*STAR), 14 Medical Drive, Singapore 117599, Singapore

[3] Yong Loo Lin School of Medicine, National University of Singapore (NUS), Singapore 117597, Singapore

*****Correspondence: Jayantha Gunaratne (jayanthag@imcb.a-star.edu.sg)

Abstract

Understanding how fermentation enhances the nutritional characteristics of legumes is essential for maximizing their potential as sustainable protein sources. This study investigates proteomics differences between soybeans and tempeh, a fermented soy product, aiming to elucidate the mechanisms that contribute to the differences in protein nutrients. Deep proteomics analysis of commercial soybean and tempeh suggest that *Rhizopus* fermentation promote proteolysis, particularly targeting Asn and Phe, and thus increased neo-peptides generation in tempeh samples. Data supports that the absence of proteinase and protease inhibitors during fermentation promote the global proteolysis effects. *In vitro* gastrointestinal digestion of tempeh demonstrated enhanced digestibility, characterized by shorter peptide chains in the digesta. Peptidomics profiling identified digest-resistant domains and peptidyl motifs, offering insights for further digestion enhancement. These findings highlight tempeh's superiority as a plant-based protein source, emphasizing its nutritional advantages through reduced antinutritional factors and improved protein bioavailability, thereby supporting its role in a healthy diet.

Keywords: Fermentation, Proteolysis, Proteolytic, Tempeh, Soy, In vitro digestion

1. Introduction

Understanding and enhancing the nutritional quality of plant-based proteins is increasingly important given the global shift toward plant-based diets for health, environmental, and ethical reasons (Gephart et al., 2016; Willett et al., 2019). Improving protein digestibility and nutrient bioavailability in legumes and other plant sources can have significant implications for nutrition science and public health, particularly in regions where plant-based diets are predominant (Basargekar et al., 2021).

Fermentation is a traditional food processing technique widely utilized across cultures to enhance the nutritional value, digestibility, and shelf life of various foods by harnessing the metabolic activities of microorganisms. It has been established that fermentation processes lead to significant protein degradation and hydrolysis, resulting in the formation of smaller peptides and free amino acids, which contribute to nutritional enhancement (de Reu et al., 1995). Despite its long-standing use, the precise molecular mechanisms underlying the nutritional enhancements provided by fermentation, such as the enzymes involved and its hydrolysis target, remain insufficiently understood.

Legumes, particularly soybeans (*Glycine max*), are esteemed for their high protein content and rich array of essential nutrients (Grdeń & Jakubczyk, 2023). However, their full nutritional potential is often hindered by antinutritional factors such as phytic acid and complex protein structures that impede nutrient absorption and protein digestibility (Samtiya et al., 2020). These limitations pose challenges for individuals relying on plant-based proteins to meet their nutritional needs.

A traditional fermented soybean product originating from Indonesia—tempeh, serves as an excellent model to study the effects of fermentation on protein nutrients. Produced through the fermentation of soybeans with fungi from the genus *Rhizopus*, tempeh has gained recognition as a plant-based meat alternative with complete protein sources (Mayer Labba et al., 2022), due to its improved nutrient profile and availability resulting from fermentation-mediated biochemical transformations (Ahnan-Winarno et al., 2021). These transformations reduce antinutrients, enhance the bioavailability of micronutrients—such as riboflavin, vitamin B6, nicotinic acid, and pantothenic acid—and minerals like magnesium, iron, and manganese (Sanke et al., 1970), and breaks down complex proteins into simpler, more digestible forms.

In this study, we aimed to unlock the potential health benefits of tempeh using advanced proteomics and peptidomics technology. Our objectives were to (i) profile the proteolytic fragments in tempeh and to (ii) assess the differences in nutrients availability between soy

and tempeh. These findings are pivotal in realizing the full potential of fermented legumes as a staple plant-based protein source.

2. Methods and materials

2.1. Soy and tempeh protein extraction

Protein of soybeans and tempeh were extracted from five different commercially available sources. Whole proteome of the soybeans (seeds of *Glycine Max*) and tempeh were extracted with the urea-based extraction buffer according to the following procedures.

Prior to start, soybeans were sterilized with 70 % of ethanol and washed thrice with distilled water. Around one gram of soybeans or tempeh were ground using a pre-chilled mortar and pestle into fine powder in liquid nitrogen. Around 100 mg of the fine powder was added in a 1.5 mL tubes and subsequently defatted with hexane for a total of three times. Defatting was performed by adding 1 mL of hexane into the tubes containing the ground soy or tempeh powder, rotated end-to-end for 15 minutes, and then removed. At the last defatting steps, a centrifugation step was performed at 11,000 g for five minutes to remove all remaining hexane.

Pellet was air-dried prior to resuspension in two pellet-volumes of extraction buffer, which contains 9 M urea, 1 % NaDoc (w/w), and 0.05 % (v/v) Benzonase in 20 mM Hepes (pH 8) buffer. Protein extracts were incubated for 20 minutes with intermittent vortex of 15 seconds at every 5-minutes intervals. The tubes were then centrifuged for 15 minutes at maximum speed or 11,000 g. The supernatant layer, which contains protein extracts, was carefully transferred into a new microcentrifuge tube. The protein extracts were diluted fifty times before protein concentration determination using the 660 assays (Thermo Scientific). Protein extracts for SDS-PAGE, gel-based proteomics were prepared using this method.

2.2. Sodium dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and gel proteomics

Extracted proteins were separated on a 10% SDS-PAGE gel (1 mm) and stained with Colloidal Blue Staining Kit (#LC6025, Invitrogen™). Each gel lane was then cut into eleven gel pieces. Gel pieces were reduced, alkylated and tryptic digested in gel using the standard procedures described previously (Shevchenko et al., 2006). Peptides from each gel piece were vacuum dried with centrifugal evaporator (CentriVap®, Labconco), reconstituted in 20 µL of 0.1 % FA prior to LC-MS/MS acquisition.

2.3. LC-MS/MS acquisition for proteomics

Vacuum-dried samples were reconstituted in 0.1 % formic acid. A total of 5 µL of peptides from each gel piece were loaded onto a PepMap™ 100 C18 trap column (#164946, 5µm, 100 Å, 100 µm x 20 mm) and a PepMap™ RSLC C18 analytical column (#164570, 2um, 100 Å, 75µm x 500 mm) using Easy nLC1000 (Themo Scientific). The mobile phase A comprised of 0.1 % formic acid in water and LC mobile phase B comprised of 0.1 % formic acid in 95 % acetonitrile. Tryptic peptides were separated under a 3-hour gradient of 8–38 % mobile phase B at 200 nL/min into Easy-spray source™ and QExactive Plus mass spectrometer (Thermo Scientific) operating in positive ion mode. Data acquisition was operated in data-dependent scanning mode, with one full MS scan followed by 20 MS2 scans of the most abundant ions, with 10 seconds of dynamic exclusion enabled. Full scan MS spectral (MS1; m/z 310 – 2000) was acquired with a resolution of 70,000 at an AGC target of 3e6 with a maximum injection time of 10 ms. Each MS2 was sequentially isolated to an AGC target value of 5e4 with resolution of 17,500 and maximum injection time of 50 ms and fragmented using HCD collision energy of 30 to acquire MS2 spectral at mass range m/z 200 – 2000 at isolation window of 2 m/z.

2.4. In vitro digestion

In-vitro digestion was performed following the standardized INFOGEST protocol (Brodkorb et al., 2019). Electrolyte stock solutions (1.25x concentration) were prepared for each phase of digestion, and enzyme activities were verified prior to their use. Soybean were soaked in water overnight and cooked in boiling water for 1 hour while tempeh was not soaked but cooked in a similar manner. All food samples were then let cool at room temperature and finely homogenized prior to the digestion. The digestion was conducted in an incubator set at 37°C (HettCube 200, Hettich, Tuttlingen, Germany) with continuous rotation at 45 rpm (Rotator SB Stuart, Huberlab, Aesch, Switzerland). In the oral phase, 5 mL of simulated salivary fluid (SSF) was added to 5 g of each sample, and 5 grams of either soy or tempeh was used in separate 50 mL centrifuge tubes. These tubes were placed in the incubator and gently shaken for 2 minutes to simulate the agitation of food in the mouth. Amylase were added during the oral phase according to INFOGEST 2.0. For the gastric phase, simulated gastric fluid (SGF) was added to the oral phase mixture to achieve a final volume ratio of 1:1 (v:v), and the pH was adjusted to 3.0. Porcine pepsin was added to reach a final activity of 2000 U/mL digesta (P7012, Sigma-Aldrich, Missouri, USA), and the mixture was incubated for 2 hours at 37°C. Gastric lipase was not added, as protein digestion was the primary focus.

In the intestinal phase, 20 mL of simulated intestinal fluid (SIF) was added to the gastric phase mixture, and the pH was adjusted to 7.0. Porcine bile salts (B8631, Sigma-Aldrich, Missouri, USA) were added to a concentration of 2.5 mM, and porcine pancreatin was added to achieve 100 U trypsin activity/mL digesta (P7545, Sigma-Aldrich, Missouri, USA). The samples were then incubated for an additional 2 hours at 37°C to simulate digestion in the small intestine. After 120 minutes, digestion was halted by adding Pefabloc® SC to a final concentration of 5 mM. The digesta were centrifuged at 13,000 x g for 15 minutes at 4°C and subsequently stored at -80°C for future analysis.

2.5. LC-MS/MS acquisition for digestomes peptidomics

Peptidomics analyses were conducted on the digesta obtained from simulated gastrointestinal digestion to characterize the peptide profiles and elucidate the protein degradation patterns. Digesta were diluted 20x in 0.1 % formic acid, centrifuged at 20,000 x g for 10 minutes at room temperature. Supernatants were transferred onto preconditioned C18 stage tips for desalting. Eluted peptides were vacuum dried with centrifugal evaporator (CentriVap®, Labconco), reconstituted in 0.1 % formic acid. A total of 1ug peptides, measured by nanodrop with 280nm absorbance, were onto a PepMap™ 100 C18 trap column (#164946, 5µm, 100 Å, 100 µm x 20 mm) and a PepMap™ RSLC C18 analytical column (#164570, 2um, 100 Å, 75µm x 500 mm) using Easy nLC1200 (Thermo Scientific). Tryptic peptides were separated under an 80-minute gradient of 5–40 % mobile phase B at 300 nL/min into Easy-spray source™ and Orbitrap Exploris 480 mass spectrometer (Thermo Scientific) operating in positive ion mode. Data acquisition was operated similar to in data-dependent scanning mode, with one full MS scan followed by 40 MS2 scans of the for the instrument settings in section 2most abundant ions, with automatic dynamic exclusion enabled. Full scan MS spectral (MS1; m/z 350 – 1500) was acquired with a resolution of 120,000 at 300% target AGC. Each dependent scan was sequentially isolated at isolation window of 1.6 m/z, to a normalized AGC target value of 50% with resolution of 15,000 and maximum injection time of 50 ms and fragmented using HCD collision energy of 30.

2.6. Data processing and analysis

Data was searched with MSFraggers (version 3.5) using FragPipe GUI V18.0 (Kong et al., 2017) against proteome database of *Glycine max* (74,863 protein entries, downloaded 20th January 2021 from Uniprot, UP000008827) and *Rhizopus oligosporus* (26 protein entries, downloaded 20th January 2021 from UniProtKB database, all protein, with taxonomy ID: 4847). Default parameters for a non-specific search were employed, with a defined precursor tolerance of ±10 ppm, and fragment tolerance of 0.06 Da, no enzyme specificity

for database digest. Cysteine carbamidomethylation were set as fixed modifications, while methionine oxidation and acetylation of protein N-termini were set as variable modifications.

Raw files were searched with decoy and built-in contaminants database, with PSM FDR estimation 0.01% allowed using Percolator. PSM, peptide, and protein level output tables files were used for visualization of the results using GraphPad Prism 9.0.

GO term of soybean proteins were extracted from Phytozome database (Goodstein et al., 2012) (JGI ID: Phytozome-109, downloaded 21st October 2022). The peptides identified after *in vitro* digestion are displayed using the peptide profile tool from the Peptigram web application (Manguy et al., 2017).

3. Result

3.1. Whole proteome profiling reveals a global proteolytic event in soybean proteins

Tempeh fermentation is a complex process involving the transformation of legumes, predominantly soy, by fungi from the family *Mucoraceae*, *Rhizopus* genus, into a palatable white cake shape tempeh. This process comprises various steps, including bean hydration (soaking), cooking, fermentation (inoculation with *Rhizopus spp.*) and caking (Ahnan-Winarno et al., 2021). We hypothesized that fermentation led to proteins breakdown into protein hydrolysates consisting of polypeptides, dipeptides, and amino acids, thereby increasing its digestibility in humans (Nout & Kiers, 2005).

Hence, we set up to monitor the biochemical transformations in tempeh's food protein composition at the molecular level. Our experimental workflow is summarized in **Figure 1A**. Our proteomics strategy focused on interpreting proteolytic events, cleavage sites, and the resultant peptides fragments during endogenous protein breakdown.

To understand the molecular basis of the protein hydrolysis, we first evaluate the degree of proteolytic cleavage of the soy proteins during the fermentation process. To assess proteome profile of soy and tempeh unbiased, we obtained five different commercially available sources of soybean and tempeh from the local market and extracted the protein lysates. As shown in **Figure 1B**, SDS-PAGE analysis of protein lysates from soy and tempeh revealed that the soy proteome preserves its integrity, as evidenced by good resolution of stained protein bands, whereas tempeh protein exhibited a general smear and dispersed pattern at a lower molecular weight (**Figure 1B**). This proteolytic pattern suggests that during tempeh fermentation, a global and extensive proteolytic event, rather than targeted protein cleavage, is responsible for this breakdown of intact proteins. Additionally, we observed noticeable differences in gel patterns among same sample types. However,

identifying these differences was not the primary focus of the current study. Nevertheless, this sample diversity provides a good foundation for comparison, enabling us to distinguish and identify the genuine effects of fermentation.

Given the extensive proteolysis observed, we sought to characterize the proteolytic patterns to identify endogenous hydrolysed peptide fragments generated during fermentation. A gel proteomics cutting each lane into 11 gel pieces were performed. A bottom-up proteomics strategy was used in combination of a non-specific enzyme search strategy in our database searches against combined proteome databases of *Glycine max* (soybean) and *Rhizopus oligosporus*, to ensure comprehensive analysis of hydrolysed neo-peptides including those containing non-tryptic peptidyl termini due to fungal fermentation.

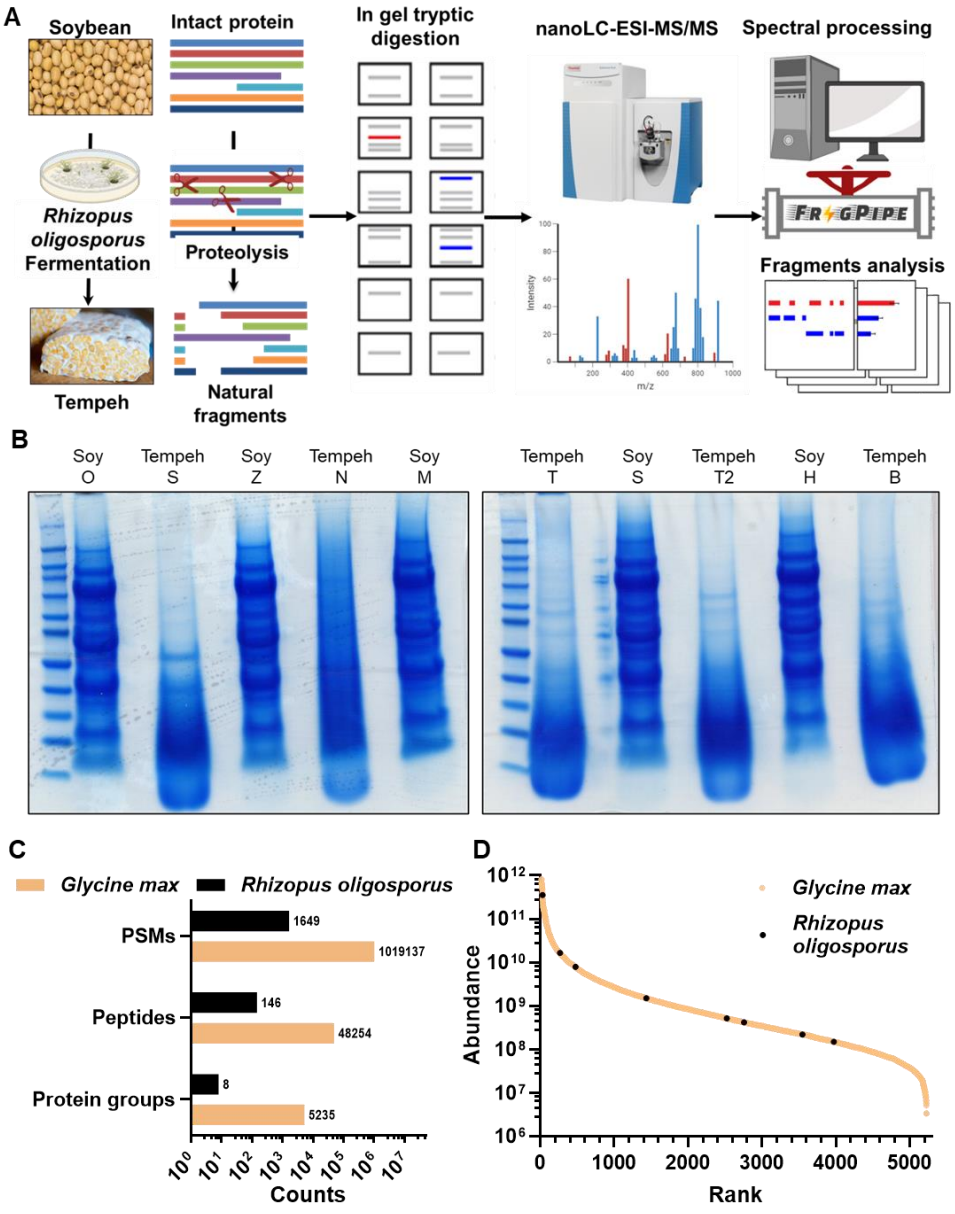

Figure 1. Tempeh fermentation induced global protein hydrolysis of intact soy protein into protein hydrolysates

(A) Illustration of the proteolysis events during tempeh fermentation mediated by *Rhizopus oligosporus* followed by a diagram of the gel-based bottom-up proteomics strategy used to interpret these proteolytic events at the molecular level. Protein lysates from soy and tempeh were separated by SDS-PAGE gel electrophoresis followed by in-gel digestion and liquid chromatography – tandem mass spectrometry (LC–MS/MS) analysis as described in the methods section. The MS raw spectral were search against the combined proteome databases of the soybean (*Glycine max*) and the organism involved in fermentation (*Rhizopus oligosporus*) to capture both the soybean proteome and *Rhizopus* proteome changes. **(B)** SDS-PAGE analysis comparing five different commercially available soy protein lysates and tempeh protein hydrolysates. **(C)** Bar graph depicts the total numbers of proteins, peptides, and PSMs from *Glycine max* (brown) and *Rhizopus oligosporus* (black) discovered across all ten samples (five soybean and five tempeh) in the bottom-up proteomics experiment, providing comprehensive overview of the proteomic landscape. **(D)** An abundance ranking graph depicts the distribution of *Glycine max* (brown) and *Rhizopus oligosporus* (black) proteins. Proteins were ranked by its overall intensity detected by mass spectrometry across all ten samples, which reflected the protein abundance.

Our proteomics analysis identified 48,254 peptides corresponding to 5,235 protein groups from the soy origins and 146 peptides corresponding to 8 *Rhizopus* protein groups across all samples (**Figure 1C**). It is noteworthy that the proteome library for *Rhizopus oligosporus* in the Uniprot database is limited, with just 26 protein entries available, as opposed to the soy proteome that contains 74,863 protein entries. The abundance rank chart, as depicted in **Figure 1D**, shows the protein rank of proteins from the *Glycine max* and *Rhizopus oligosporus* origins, respectively.

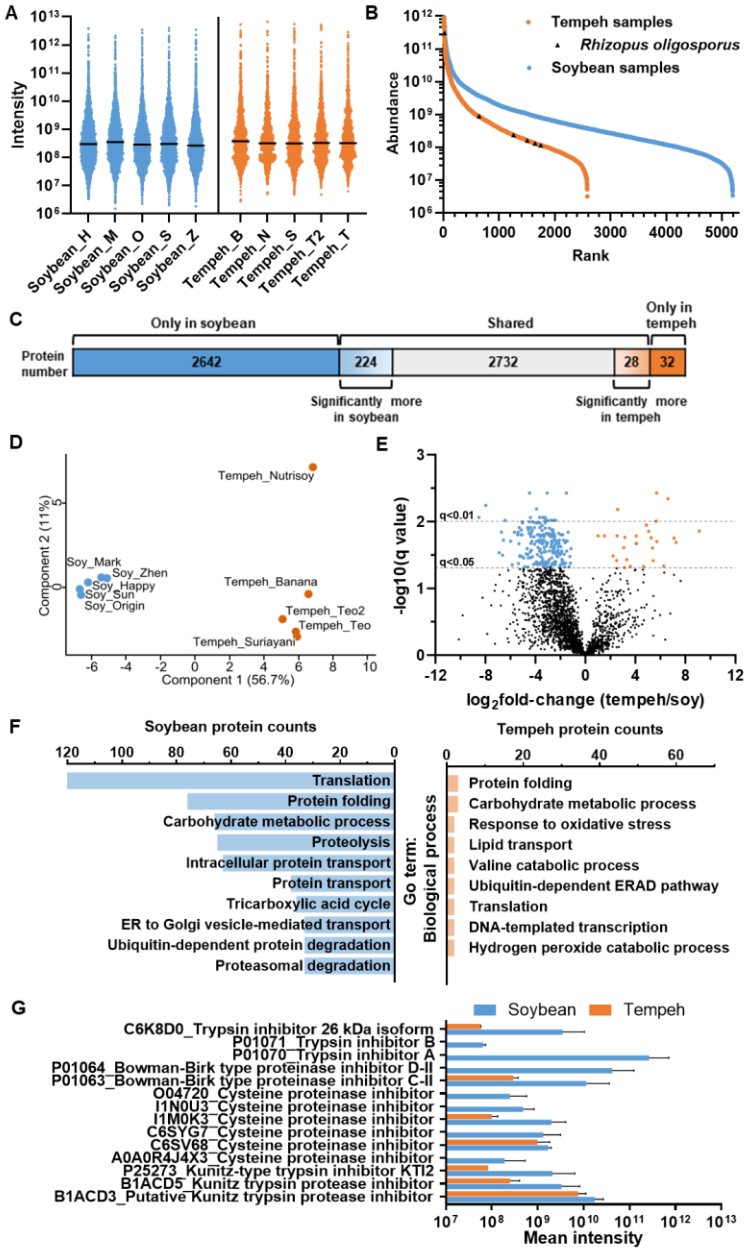

Figure 2. Differential expressed protein comparing soy and tempeh. **(A)** Intensity of proteome across all samples**. (B)** An abundance ranking graph based on intensity-based quantitation values depicts the distribution of *Glycine max* (round dots) and *Rhizopus oligosporus* (triangular dots) in both soybean and tempeh samples. **(C)** Bars indicate a Venn diagram of protein expression, boxes on both side shows protein that are either detected only in soybean, or in tempeh samples. Gray scalebars indicate genes that existed in both soybean and tempeh samples (shared) and among those the significantly regulated ($p<0.05$) is indicated with coloured gradient box. **(D)** A Principal Component Analysis (PCA) illustrating the multivariate relationships between the protein composition of soy and tempeh. **(E)** Volcano plot of differential expressed protein in both soybean and tempeh. Coloured dots indicate significantly differential proteins. **(F)** A Gene Ontology (GO) analysis of differentially expressed proteins in both soybean and tempeh. **(G)** Intensity of commonly known protease and proteinase inhibitors from the soy origins in both soybean and tempeh samples. Mean protein intensity across all five samples were presented.

Bottom-up proteomics analysis revealed clear global proteolytic differences between soybean and tempeh proteins. Proteome profiling of soy and tempeh indicate that the distribution of the proteome abundance comparing soy and tempeh is similar on the log scale (**Figure 2A**), with more proteins identified from the soybean samples compared to the tempeh sample (**Figure 2B, C**). **Figure 2C** shows the Venn diagram that depicts the protein found in soy and tempeh samples. 2642 proteins (~50% of total protein groups) were specifically detected in soy whereas 32 proteins were specifically identified in the tempeh. Among the proteins shared by both soybean and tempeh (2984 proteins), 8.4% showed a statistically significant change in abundance ($p<0.05$), as indicated by coloured gradient bars (**Figure 2C**).

The principal component analysis of the protein composition (**Figure 2D**) supported that there's fundamental differences in the proteome profile between soy and tempeh products. Looking into the protein composition difference between soy's and tempeh's commercial samples, 224 proteins are found to have significantly higher abundance in soy whereas 28 proteins are significantly more abundant in tempeh, as shown in the volcano plot (**Supplementary Table 1, Figure 2E**). These results align with our expectation that certain proteins would be significantly reduced or absent in tempeh due to fermentation-induced proteolysis. In addition, the hydrolysis of high-abundance proteins during fermentation appears to increase the sensitivity of proteomic detection for low abundant proteins in the tempeh.

3.2. Fermentation process led to allergens and antinutritional factors breakdown

Since soy is known as a source of allergens, we wondered if proteolytic activity in tempeh fermentation might reduce the allergen sources in soy. A search for commonly known allergenic fragments shows that tempeh has significantly less allergenic fragments compared to soy, as depicted as **Supplementary Figure 1**, shown as intensity of tempeh fragments in each soy-derived gene.

Additionally, we explored the roles of proteins that were specifically present or significantly more abundant in either soy or tempeh. Mapping these genes to the Gene Ontology (GO) terms for biological processes (**Figure 2F**) revealed that proteins related to protein synthesis and regulation, such as protein translation, folding, transport, proteolysis, and degradation, were specific to soybean but absent or lower in tempeh. This suggests that the fermentation process involving *Rhizopus* spp. influences protein synthesis-related biological processes in soybean. Other biological processes that decreased in tempeh included carbohydrate metabolism and the tricarboxylic acid (TCA) cycle.

Soybean is known to contain proteinase inhibitors help regulate the activity of endogenous proteases during seed development and germination (Hwang et al., 1978; Vorster et al., 2023). We next asked the impact of tempeh fermentation on the abundance of protease inhibitors. Since there is no specific Gene Ontology (GO) term annotated for soybean protease inhibitors, we manually curated a list of 14 protease and proteinase inhibitors based on a keyword search for "inhibitors" in the gene descriptions. This list includes families such as cysteine proteinase inhibitors, Kunitz-type trypsin protease inhibitors, and Bowman-Birk type proteinase inhibitors. **Figure 2G** depicts the abundance of protease/proteinase inhibitors from both soybean and tempeh samples. We discovered that, from the 14 inhibitors, 7 were absent in all five commercial tempeh samples, where the remaining 7 were found in both soybeans and tempeh samples. This suggests that protease and proteinase inhibitors in the soybean were susceptible to the fermentation induced proteolysis, which likely contributed to accelerated process of global proteolysis.

3.3. Fermentation increases neo-peptides generation

Knowing that tempeh fermentation led to proteolysis (**Figure 1B**), we next sought out to investigate the type of proteolysis occurred during fermentation. Moving from protein to peptides analysis, **Figure 3A** depict the peptides length distribution (ranging from 7–25 amino acids) of soybean and tempeh samples after bottom-up proteomics workflow which include a trypsin digestion. As predicted, tempeh samples yielded higher abundance of lower molecular weight oligomer compared to the soybean samples, indicating that the peptides were likely to contain a cleavage site resulted from the proteolysis, which truncated the peptides into a shorter peptide, such as 9-mer to 13-mer. Overall, this suggests active proteolysis in the tempeh samples, likely driven by proteases by *Rhizopus spp.*

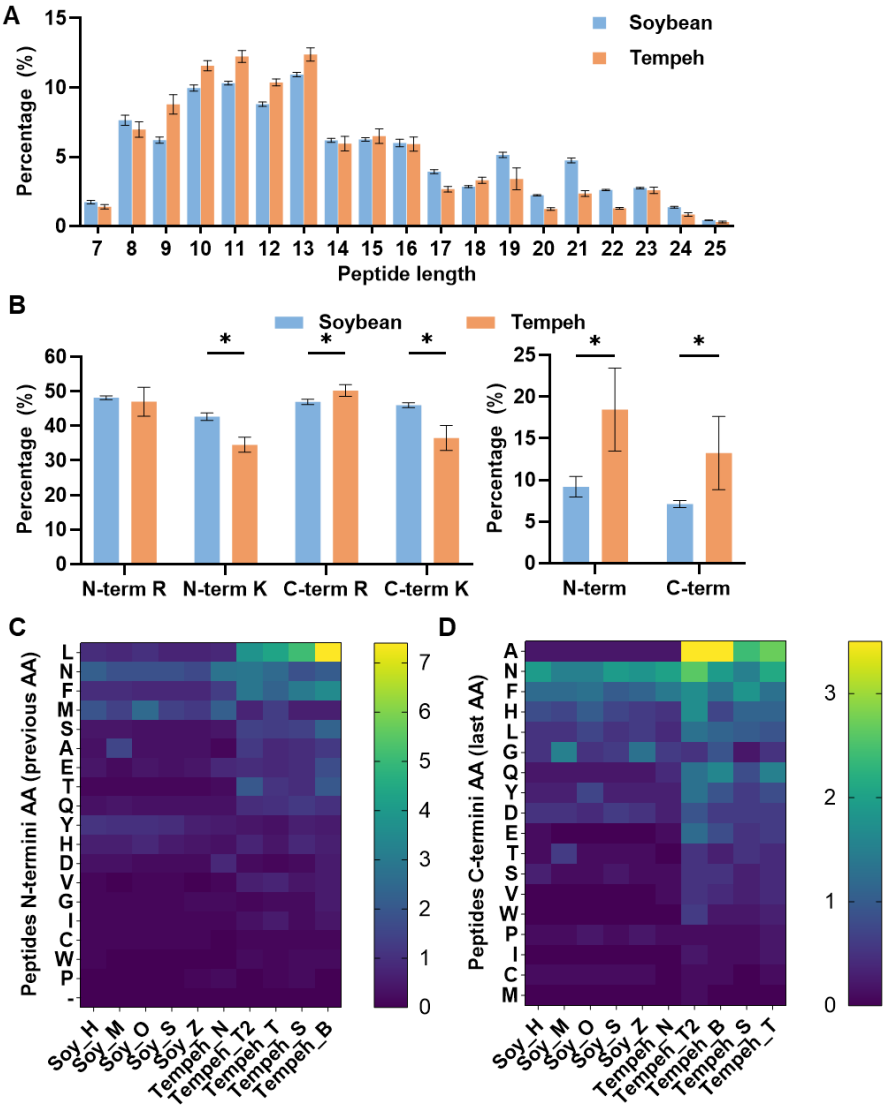

Figure 3. Peptides-centric analysis reveal more neo peptides fragments generation in tempeh. (A) The top panel displays the distribution of peptides abundance in both soybean and tempeh samples in peptides length spanning from 7-25 AAs. **(B)** Bar graphs show percentage of the peptides containing either tryptic cut sites (carboxyl termini of arginine and lysine), and non-tryptic cut sites, at peptides N-term and C-term. **(C-D)** Two heatmaps depicting the further breakdown of the peptides non-tryptic cleavage sites carboxyl-termini of the peptides N-termini **(C)** and peptides C-termini **(D).** The heatmaps are color-coded based on the frequency of each non-tryptic AA appeared in the peptides N or C-terminal in percentage (%), ranging from low (purple) to high (yellow). Each row represents a samples wither from soy or tempeh.

To further validate on our hypothesis, we performed peptides-termini analysis to investigate neo fragments generated by the proteases of *Rhizopus oligosporus*. Unlike trypsin, which cleaves peptides specifically at the carboxyl side of lysine and arginine residues, fermentation process exhibits a broader substrate specificity, leading to the formation of peptides with non-tryptic termini. By allowing a nonspecific cleavage site in our database

searches, we identified a significant number of peptides with neo-N or neo-C termini that were not present in the unfermented soybean samples. As shown in **Figure 3B**, majority of peptides (>~80%) contain a tryptic site, whereas 5-20% of peptides contain a non-tryptic cleavages site, across all samples.

Summing all the non-tryptic cut sites (all peptides' termini except when contain arginine and lysine), tempeh samples contain about two times more of peptides carrying non-tryptic cleavage sites compared to the soy samples (**Figure 3B,** right panel**;** 9.18% vs. 18.44% for N termini, 7.1% vs. 13.24% in C termini), confirming our hypothesis that these neo-terminal fragments are indicative of the extensive proteolytic activity exerted by fungal enzymes during fermentation. In addition, it is noteworthy that the tempeh and soybean samples contain similar number of arginine cleavage sites (~47–48%), whereas tempeh samples contain less lysine tryptic cleavage site (~34.5–36.5%) compared to soybean samples (42.746%), suggesting that the sequence containing arginine are likely more resistant to further proteolysis.

Next, we further analyse the non-tryptic cut sites to see whether there's an over-represented cleavages site in these neo generated fragments. **Figure 3C-D** depicted the AA lies in carboxyl-termini of non-tryptic peptides cleavage. Notably, we observed a trend of higher Phe (F) and Asn (N) cleavage sites in both C-and N-terminal of peptides of tempeh samples. In addition, higher Leu (L) cleavage activity was observed in peptides N-terminal of tempeh, when compared to the soybean. These observations collectively support the notion that *Rhizopus spp.* contributes to a more diverse peptides cleavage through generating non-tryptic cleavage site.

To summarize, our data underscores the global proteolytic changes in soy proteins, resulting in the breakdown of high-molecular-weight proteins into smaller, more digestible peptides, as evidenced in **Figure 3A**; while at the same time, highlights the specificity of preferred cleavage sites, as demonstrated in **Figure 3C-D**. This extensive proteolysis might enhance the nutritional profile of tempeh by increasing the availability of essential amino acids but also contributes to its improved digestibility.

3.4. Fermentation improves protein digestibility

Next, we further investigated the significance of proteome differences in relation to human digestion, aiming to understand whether these differences could potentially translate to mechanisms of food nutrient absorption. The Infogest protocol provides a standardized method for conducting *in vitro* digestion studies, ensuring consistency and comparability across different research initiatives. As illustrated in **Figure 4A,** the digesta undergoes

peptidomics analysis using LC-MS/MS to comprehensively assess the digestibility of food peptides. This advanced analytical technique allows for the identification and quantification of a wide range of peptides present in the digesta, providing insights into dynamic interactions between food peptides and the gastrointestinal environment.

Figure 4B indicate that fermented-soy, tempeh results in better digestibility, as it produces significantly shorter digested peptides, as evidenced by the higher abundance of peptides with length ranging 8 – 11, and lower abundance of tempeh peptides with length > 13, suggesting heightened susceptibility to digestive protease action. Next, we looked into the sequence of peptidyl termini to understand whether there's preference in digestive enzymes. **Figure 4C-D** depict the sequence motifs derived from peptidyl termini in the gastrointestinal tract, which reveals distinct substrate specificities among the represented enzymes. The N-terminal shows an overrepresentation of acidic amino acids, specifically glutamic acid (G) and aspartic acid (D), at positions near the cleavage sites. This pattern suggests resistant of digestion due to negatively charged environments or structure caused by acidic residues. In contrast, the C-terminal of sequence motif demonstrates a stronger preference for basic residues, notably for soy, arginine (R) and Histidine (H) at the position immediately adjacent to the cleavage site. In general, the profile exhibits a mix of polar and charged residues across the spectrum, suggesting a broad enzymatic activity that targets a variety of peptide bonds. Noteworthy, proline (P) was overrepresented at the second last peptidyl bond (-2 position), suggesting the propyl-x peptide bonds was not accessible by digestive enzymes, partly due to the aromatic ring of proline. Taken together, these findings imply that the partial hydrolysis of soy proteins by *R. oligosporus* during fermentation renders them more accessible to digestive enzymes, thereby enhancing nutritional bioavailability.

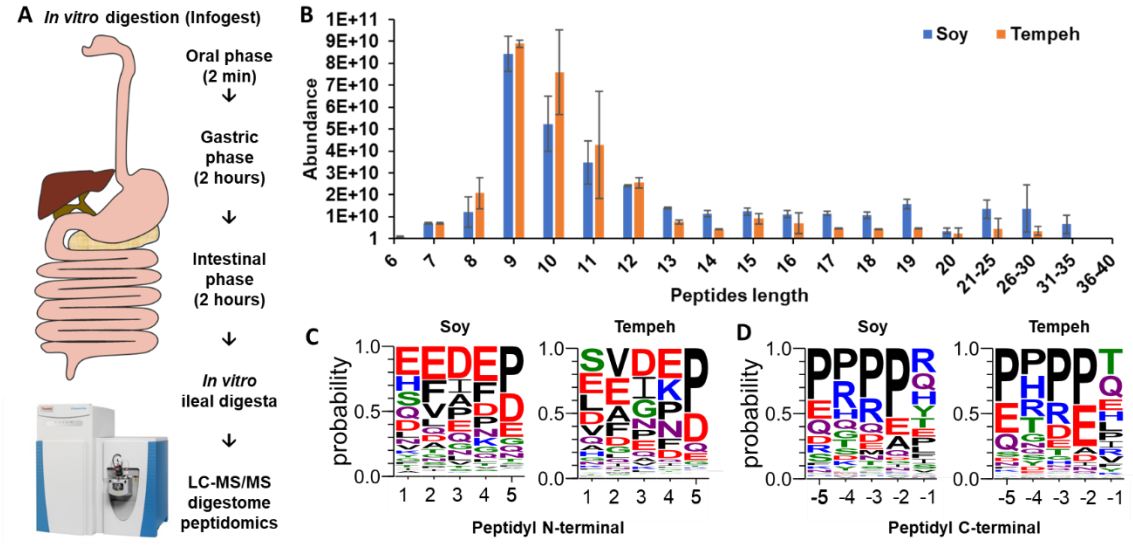

Figure 4. Digest-peptidome reveals improved digestibility of tempeh. (A) Graphical illustration of the *in vitro* human gastrointestinal system and peptidomics analysis. **(B)** Bar graph depicts the lengths of peptides present in the digestome, providing insights into their relative digestibility. **(C-D)** Sequence motif analysis reveal the probability of over-represented amino acids found at the termini of peptides. Each position in the sequence logos (from left to right for N termini or right to left for C termini) corresponds to a specific residue position relative to the cleavage site.

Finally, to pinpoint whether certain protein domains or motifs remain resistant after both fermentation and in vitro digestion, we mapped the location of identified peptides across the most abundant soybean proteins. We select the top 6 soy protein with highest abundance to further understand the digestibility of these proteins. Each horizontal line in the figure represents a distinct sample, and vertical bars indicate where peptides were detected along the amino acid sequence. Digestive pattern in **Figure 5A-F** suggests that despite fermentation, there remains specific domain or motif in the protein remained not broken down. However, comparing the mapped peptides region of soybean and tempeh, it suggests that the fermentation in tempeh partly breakdown the proteins, which resulted in fewer digestion resistant protein domain mapped in tempeh compared to the soybean samples. In all, our *in vitro* digestion demonstrated that the tempeh has better digestibility, as evidenced by shorter peptides length and less resistant protein domain, compared to the soybean samples. These results further confirm that fermentation alleviates digestion bottlenecks, improving overall protein digestibility relative to unfermented soybean.

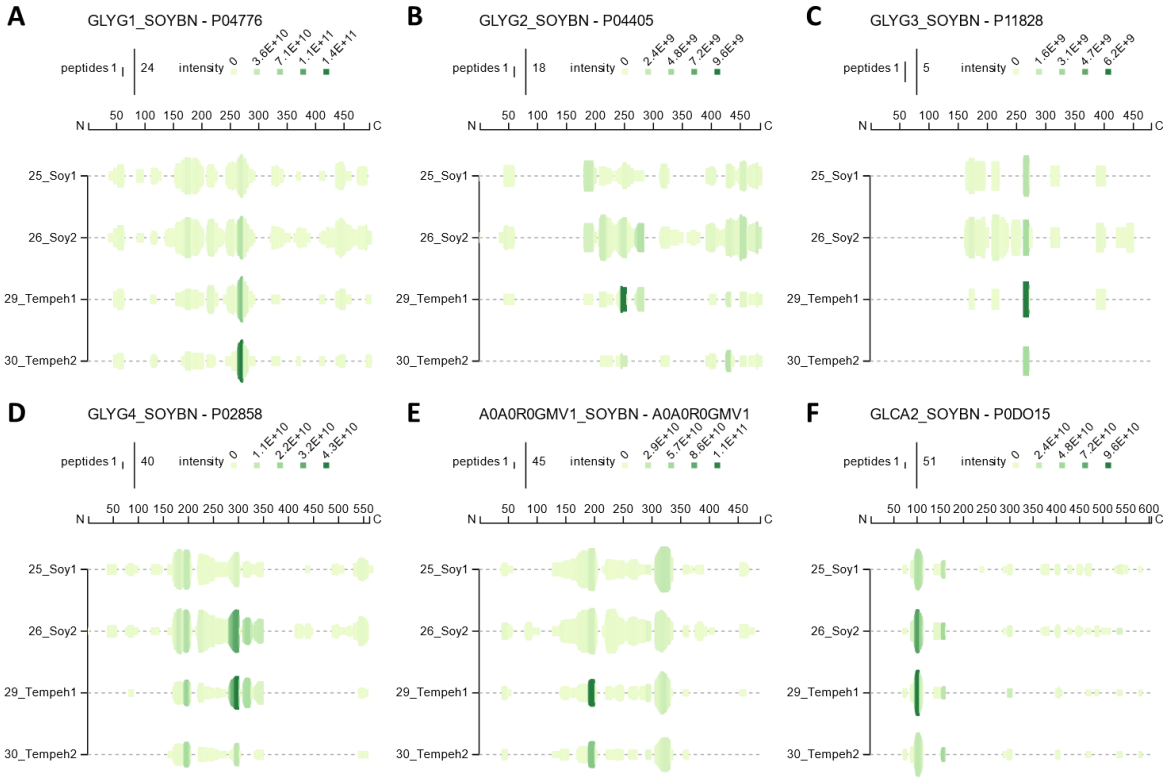

Figure 5. Peptide resistance to digestion mapped across most abundant protein sequences. This figure illustrates the distribution of peptides at the end of *in vitro* digestion across the TOP6 protein identified. Each horizontal line corresponds to a distinct sample, while the vertical bars indicate identified peptides present in specific location of a protein, with colour scale intensity demonstrating its overlapped and cumulative relative abundance within a protein species.

4. Discussion

To the best of our knowledge, this is the first study to profile protein hydrolysates in tempeh. One of the key findings is the identification of over 48,000 peptides derived from soybean proteins.

A central finding of our study is the extensive proteolysis of soy proteins during tempeh fermentation generate protein hydrolysates, and liberation of neo-peptides during this process. Deep proteome profiling in this study highlights that tempeh fermentation significantly increase proteolytic cleavages. The PCA plot indicate that fermentation led to similar proteome outcomes despite variations in sources of commercial soy and tempeh, suggesting that the fermentation involving *Rhizopus spp.* plays a dominant role in shaping the final products. Data supports that the fermentation process pre-digests the proteins, making them more accessible for human gastrointestinal enzymes. This pre-hydrolysis likely contributes to the enhanced nutritional profile of tempeh, supporting its role as a high-quality protein source in plant-based diets.

4.1. Mechanisms of Fermentation and Enzymatic Activity

Fermentation is a well-established method for improving the nutritional profile of food products. The fermentation of soybeans to produce tempeh is a complex biochemical process involving hydration, cooking, inoculation, and caking. Hydration of soybeans is a critical step that activates endogenous enzymes and prepares the substrate for fermentation. Soaking the beans in water initiates the activation enzymes, which begin breaking down complex macromolecules. This activation increases the accessibility of nutrients and facilitates the action of microbial enzymes during fermentation (Ahnan-Winarno et al., 2021), which secretes a variety of extracellular enzymes, including proteases, lipases, and amylases, which further degrade macromolecules into simpler molecules (Nout & Kiers, 2005).

The lack of availability of protease entries in *Rhizopus oligosporus* databases limits the comprehensive understanding of these enzymatic processes. Nevertheless, the identification of neo-terminal peptides highlights the pivotal role of fungal proteases in transforming soy proteins during tempeh fermentation. In addition, non-tryptic cleavage sites were more

prevalent in tempeh, highlighting the broader substrate specificity of fungal enzymes compared to trypsin. At the same time, overrepresented cleavage sites at the neo-peptides such as Phe and Asn provide insights of the mechanism of action of the Rhizopus enzymes.

Another possibility explanation of the proteolysis is that it might be owing to the plant's natural protease in the fermentation process, with *Rhizopus* aiding the digestion of complex carbohydrates structure to make protein accessible. Soybeans it known to contain papain-like cysteine proteases, which are involved in regulating plant growth and stress responses (Mangena, 2020). Another potential mechanism demonstrated by our findings is that fermentation effectively reduces antinutritional factors in soybeans, particularly protease and proteinase inhibitors such as the Bowman-Birk Protease Inhibitor (Martinez et al., 2019), which is known to inhibit trypsin and chymotrypsin inhibitor, two important digestive enzymes.

4.2. Nutritional benefits of soybean fermentation

Another significant finding is the reduced allergenicity of tempeh compared to soy. This reduction in allergenic fragments in tempeh can be crucial for individuals with soy allergies. Tempeh may offer a safer dietary option for those with soy allergies, enabling them to enjoy the nutritional benefits of soy without the risk of allergenic reactions. Further studies, particularly in clinical settings, can explore this aspect of tempeh's safety and tolerability.

Next, improved digestibility of tempeh proteins compared to unfermented soybeans is another significant outcome of this research. The peptide length distribution of the *in vitro* digesta showed that tempeh contained a higher abundance of shorter peptides compared to intact soy proteins, which likely contributed to improved protein accessibility. Our data supports the application of tempeh fermentation in animal nutrition, particularly in poultry and swine diets, in which protease supplements are often included to improve the digestibility of soybean meal to enhance the availability of amino acids and improve overall nutrient absorption (Kiarie et al., 2020; Mohammadigheisar et al., 2021).

Previous studies have highlighted the health benefits of consuming fermented foods (Fujita & Yamagami, 2001; Kim et al., 2008; Kwon et al., 2010; Taniguchi et al., 2008), such as tempeh over unfermented counterparts. Studies comparing effects of soy- and tempeh-feeding in diabetic rat models reveal that rats fed tempeh perform better on the oral glucose tolerance test than rats fed soy (Huang et al., 2018; Park et al., 2012), suggesting that fermented soy contains additional beneficial components resulting from the fermentation process. One such component is protein hydrolysates (Lau & Dunn, 2018), which consist of a mixture of amino acids and peptides resulting from soy protein hydrolysis during

fermentation. Our digesta profiling not only provides a readily available database to look into potential bioactive components resulted from the fermentation, but also revealed the digestion “roadblocks” or bottleneck, providing a target to improve on.

4.3. Limitations of the study

A notable limitation of this study is the use of commercially sourced soybean and tempeh products, which restricts our visibility into the specific cultivars of soybeans and *Rhizopus* strains used in the fermentation process. Nevertheless, principal components analysis of the proteome (Figure 2D) suggests that tempeh underwent similar protein hydrolysis despite different brands of commercial tempeh. To address these limitations in future research, it would be beneficial to source soybeans of known cultivars and produce tempeh with well-characterized Rhizopus strains in lab, providing greater control over experimental variables and allowing for more mechanistic analysis of the fermentation process and its effects on the nutritional and functional properties of tempeh.

Another significant limitation of this study is the inability to provide a detailed mechanistic explanation for the degradation of major allergens and protease inhibitors during the fermentation process. While we observed a substantial reduction or complete absence of major allergens in tempeh compared to soybeans, the exact mechanisms of this degradation remain unclear from our data. This is owing to the absence of most allergens and protease inhibitors in the tempeh samples, which precluded us from conducting a proteolytic cleavage analysis by mapping peptides.

This highlights the need for future studies employing time-course experiments. Such investigations would provide a more comprehensive understanding of how fermentation aids in reducing soybean allergens and could potentially inform strategies for producing hypoallergenic soy products.

5. Conclusion

This research offers valuable insights into the nutritional, allergenic, and proteomic characteristics of tempeh, highlighting its potential as a sustainable and nutritious plant-based protein source. The study demonstrates the significant proteome changes that occur during tempeh fermentation, particularly the extensive proteolysis of soy proteins, which leads to contribute to improved nutrient bioavailability and a potential reduction in allergenicity compared to unfermented soybeans. This proteolytic transformation significantly improves the digestibility of tempeh, facilitating the absorption of essential amino acids. This finding underscore tempeh's value as a high-quality protein profile.

Future research could delve deeper into the functional properties of these liberated peptides and explore their potential applications in functional foods and dietary supplements. Additionally, investigating consumer preferences and sensory attributes could help increase the acceptance and integration of tempeh into a wider range of diets and cuisines.

Extending these findings to other legumes and plant-based proteins, it is essential to explore similar fermentation processes that can enhance the nutritional quality and digestibility of alternative sources such as lentils, chickpeas, and peas. By understanding the proteolytic patterns of these legumes, we can develop innovative, nutrient-dense, and allergen-reduced plant-based products, supporting the diversification of plant-based protein sources and promoting sustainable dietary practices.

Acknowledgements

The authors express their gratitude to the Agency for Science, Technology and Research (A*STAR) for their support in this work.

Author contributions: CRediT

The authors' responsibilities were as follows: **Jia Yee Wu**: Investigation, Formal analysis, Writing – original draft, Writing – review & editing. **Sheena Wee:** Conceptualization, Investigation, Supervision. **Siok Ghee Ler:** Investigation. **Christiani Jeyakumar Henry**: Conceptualization, Funding acquisition, Supervision. **Jayantha Gunaratne**: Conceptualization, Supervision, Funding acquisition, Writing – review & editing. All authors have read and approved the final manuscript.

Funding sources

This research was supported by the Agency for Science, Technology and Research (A*STAR), under the Singapore Food Story R&D Programme, SFS-2 IAF-PP Future Foods: Alternative Proteins (H20H8a002). This research was also supported by the Singapore Institute of Food and Biotechnology Innovation (SIFBI) and Institute of Molecular and Cell Biology (IMCB), Agency for Science, Technology and Research (A*STAR), Singapore.

Conflicts of 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 article.

References

Ahnan-Winarno, A. D., Cordeiro, L., Winarno, F. G., Gibbons, J., & Xiao, H. (2021). Tempeh: A semicentennial review on its health benefits, fermentation, safety, processing, sustainability, and affordability. *Compr Rev Food Sci Food Saf*, *20*(2), 1717-1767. https://doi.org/10.1111/1541-4337.12710

Basargekar, P., Priyadarshini, S., Seth, S., & Ganjoo, V. (2021). Impact of socio-economic factors in reducing malnutrition among children: A comparative study of India, Bangladesh and Sri Lanka. *Asia Pacific Journal of Health Management*, *16*(3), 21-28.

Brodkorb, A., Egger, L., Alminger, M., Alvito, P., Assunção, R., Ballance, S., Bohn, T., Bourlieu-Lacanal, C., Boutrou, R., & Carrière, F. (2019). INFOGEST static in vitro simulation of gastrointestinal food digestion. *Nature protocols*, *14*(4), 991-1014.

de Reu, J. C., Marcel ten Wolde, R., de Groot, J., Nout, M. R., Rombouts, F. M., & Gruppen, H. (1995). Protein hydrolysis during soybean tempe fermentation with Rhizopus oligosporus. *Journal of Agricultural and Food Chemistry*, *43*(8), 2235-2239.

Fujita, H., & Yamagami, T. (2001). Fermented soybean-derived Touchi-extract with anti-diabetic effect via alpha-glucosidase inhibitory action in a long-term administration study with KKAy mice. *Life Sci*, *70*(2), 219-227. https://doi.org/10.1016/s0024-3205(01)01381-9

Gephart, J. A., Davis, K. F., Emery, K. A., Leach, A. M., Galloway, J. N., & Pace, M. L. (2016). The environmental cost of subsistence: Optimizing diets to minimize footprints. *Sci Total Environ*, *553*, 120-127. https://doi.org/10.1016/j.scitotenv.2016.02.050

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., & Putnam, N. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic acids research*, *40*(D1), D1178-D1186.

Grdeń, P., & Jakubczyk, A. (2023). Health benefits of legume seeds. *Journal of the Science of Food and Agriculture*, *103*(11), 5213-5220.

Huang, Y. C., Wu, B. H., Chu, Y. L., Chang, W. C., & Wu, M. C. (2018). Effects of Tempeh Fermentation with Lactobacillus plantarum and Rhizopus oligosporus on Streptozotocin-Induced Type II Diabetes Mellitus in Rats. *Nutrients*, *10*(9). https://doi.org/10.3390/nu10091143

Hwang, D. L., Yang, W.-K., Foard, D. E., & Lin, K.-T.-D. (1978). Rapid release of protease inhibitors from soybeans: Immunochemical quantitation and parallels with lectins. *Plant physiology*, *61*(1), 30-34.

Kiarie, E. G., Parenteau, I. A., Zhu, C., Ward, N. E., & Cowieson, A. J. (2020). Digestibility of amino acids, energy, and minerals in roasted full-fat soybean and expelled-extruded soybean meal fed to growing pigs without or with multienzyme supplement containing fiber-degrading enzymes, protease, and phytase. *J Anim Sci*, *98*(6). https://doi.org/10.1093/jas/skaa174

Kim, D. J., Jeong, Y. J., Kwon, J. H., Moon, K. D., Kim, H. J., Jeon, S. M., Lee, M. K., Park, Y. B., & Choi, M. S. (2008). Beneficial effect of chungkukjang on regulating blood glucose and pancreatic beta-cell functions in C75BL/KsJ-db/db mice. *J Med Food*, *11*(2), 215-223. https://doi.org/10.1089/jmf.2007.560

Kong, A. T., Leprevost, F. V., Avtonomov, D. M., Mellacheruvu, D., & Nesvizhskii, A. I. (2017). MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat Methods*, *14*(5), 513-520. https://doi.org/10.1038/nmeth.4256

Kwon, D. Y., Daily, J. W., 3rd, Kim, H. J., & Park, S. (2010). Antidiabetic effects of fermented soybean products on type 2 diabetes. *Nutr Res*, *30*(1), 1-13. https://doi.org/10.1016/j.nutres.2009.11.004

Lau, J. L., & Dunn, M. K. (2018). Therapeutic peptides: Historical perspectives, current development trends, and future directions. *Bioorg Med Chem*, *26*(10), 2700-2707. https://doi.org/10.1016/j.bmc.2017.06.052

Mangena, P. (2020). Phytocystatins and their Potential Application in the Development of Drought Tolerance Plants in Soybeans (Glycine max L.). *Protein Pept Lett*, *27*(2), 135-144. https://doi.org/10.2174/0929866526666191014125453

Manguy, J., Jehl, P., Dillon, E. T., Davey, N. E., Shields, D. C., & Holton, T. A. (2017). Peptigram: a web-based application for peptidomics data visualization. *Journal of proteome research*, *16*(2), 712-719.

Martinez, M., Gómez-Cabellos, S., Giménez, M. J., Barro, F., Diaz, I., & Diaz-Mendoza, M. (2019). Plant proteases: from key enzymes in germination to allies for fighting human gluten-related disorders. *Frontiers in plant science*, *10*, 721.

Mayer Labba, I.-C., Steinhausen, H., Almius, L., Bach Knudsen, K. E., & Sandberg, A.-S. (2022). Nutritional composition and estimated iron and zinc bioavailability of meat substitutes available on the Swedish market. *Nutrients*, *14*(19), 3903.

Mohammadigheisar, M., Shouldice, V. L., Torrey, S., Widowski, T. M., Ward, N. E., & Kiarie, E. G. (2021). Growth performance, organ attributes, nutrient and caloric utilization in broiler chickens differing in growth rates when fed a corn-soybean meal diet with multienzyme supplement containing phytase, protease and fiber degrading enzymes. *Poult Sci*, *100*(9), 101362. https://doi.org/10.1016/j.psj.2021.101362

Nout, M. J., & Kiers, J. L. (2005). Tempe fermentation, innovation and functionality: update into the third millenium. *J Appl Microbiol*, *98*(4), 789-805. https://doi.org/10.1111/j.1365-2672.2004.02471.x

Park, S., Kim, D. S., Kim, J. H., Kim, J. S., & Kim, H. J. (2012). Glyceollin-containing fermented soybeans improve glucose homeostasis in diabetic mice. *Nutrition*, *28*(2), 204-211. https://doi.org/10.1016/j.nut.2011.05.016

Samtiya, M., Aluko, R. E., & Dhewa, T. (2020). Plant food anti-nutritional factors and their reduction strategies: an overview. *Food Production, Processing and Nutrition*, *2*, 1-14.

Sanke, Y., Miyamoto, T., & Murata, K. (1970). Studies on the nutritional value of tempeh. IV. Biosynthesis of folate compounds with Rhizopus oligosporus. *J Vitaminol (Kyoto)*, *17*(2), 96-100. https://doi.org/10.5925/jnsv1954.17.96

Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V., & Mann, M. (2006). In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc*, *1*(6), 2856-2860. https://doi.org/10.1038/nprot.2006.468

Taniguchi, A., Yamanaka-Okumura, H., Nishida, Y., Yamamoto, H., Taketani, Y., & Takeda, E. (2008). Natto and viscous vegetables in a Japanese style meal suppress postprandial glucose and insulin responses. *Asia Pac J Clin Nutr*, *17*(4), 663-668. https://www.ncbi.nlm.nih.gov/pubmed/19114406

Vorster, J., Westhuizen, W. v. d., Plessis, G. d., Marais, D., Sparvoli, F., Cominelli, E., Camilli, E., Ferrari, M., Le Donne, C., & Marconi, S. (2023). In order to lower the antinutritional activity of serine protease inhibitors, we need to understand their role in seed development. *Frontiers in plant science*, *14*, 1252223.

Willett, W., Rockstrom, J., Loken, B., Springmann, M., Lang, T., Vermeulen, S., Garnett, T., Tilman, D., DeClerck, F., Wood, A., Jonell, M., Clark, M., Gordon, L. J., Fanzo, J., Hawkes, C., Zurayk, R., Rivera, J. A., De Vries, W., Majele Sibanda, L., . . . Murray, C. J. L. (2019). Food in the Anthropocene: the EAT-Lancet Commission on healthy diets from sustainable food systems. *Lancet*, *393*(10170), 447-492. https://doi.org/10.1016/S0140-6736(18)31788-4