Food Research International 186 (2024) 114348

Contents lists available at ScienceDirect

Food Research International

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

Differential effects of heating modes on the immunogenic potential of soy-derived peptides released after *in vitro* infant digestion

Jiaying Tang [a], Sjef Boeren [b], Harry J. Wichers [c,d], Kasper A. Hettinga [a,*]

[a] Food Quality & Design Group, Wageningen University & Research, Wageningen, The Netherlands
[b] Laboratory of Biochemistry, Wageningen University & Research, Wageningen, The Netherlands
[c] Wageningen Food & Biobased Research, Wageningen University & Research, Wageningen, The Netherlands
[d] Laboratory of Food Chemistry, Wageningen University and Research, Wageningen, The Netherlands

ARTICLE INFO

Keywords:
Soy protein
Heat treatment
Infant digestion
Immunogenic peptide
Peptidomics
Bioinformatics

ABSTRACT

During production of soy-based infant formula, soy protein undergoes heating processes. This study investigated the differential impact of heating modes on the immunogenic potential of peptides in soy protein digests. Wet or dry heating was applied, followed by *in vitro* gastrointestinal infant digestion. The released peptides were analyzed by LC-MS/MS. Bioinformatics tools were utilized to predict and identify potential linear B-cell and T-cell epitopes, as well as to explore cross-reactivity with other legumes. Subsequently, the peptide intensities of the same potential epitope across different experimental conditions were compared. As a result, we confirmed the previously observed enhancing effect of wet heating on infant digestion and inhibitory effect of dry heating. A total of 8,546 peptides were detected in the digests, and 6,684 peptides were with a score over 80. Among them, 29 potential T-cell epitopes and 27 potential B-cell epitopes were predicted. Cross-reactivity between soy and other legumes, including peanut, pea, chickpea, lentil, kidney bean, and lupine, was also detected. Overall, heating and digestion time could modulate the potential to trigger peptide-induced immune responses.

1. Introduction

Soy protein-based infant formula (IF) is commercially used as a substitute for dairy-based IF, as it is suitable for infants with cow's milk allergy (CMA), lactose intolerance, or those from vegan families. However, approximately 10–14 % of infants with CMA also exhibited allergic reactions to soy protein (Kattan, Cocco, & Järvinen, 2011; Klemola, Vanto, Juntunen-Backman, Kalimo, Korpela, & Varjonen, 2002; Zeiger et al., 1999). This is because people with one food allergy may be more prone to developing additional allergies due to a heightened immune system response (Hill & Spergel, 2018). Furthermore, according to WHO/IUIS Allergen Nomenclature Sub-committee (https://www.allergen.org/, accessed April 2023), eight common soy allergens involve hydrophobic protein (Gly m 1), defensin (Gly m 2), profilin (Gly m 3), pathogenesis-related protein (Gly m 4), β-conglycinin (Gly m 5; vicilin, 7S globulin), glycinin (Gly m 6; legumin, 11S globulin), seed biotinylated protein (Gly m 7), and 2S albumin (Gly m 8). Amongst them, β-conglycinin and glycinin have been extensively studied due to their association with severe soy allergies (Holzhauser et al., 2009; Ito et al., 2011). Furthermore, individuals allergic to soy may also demonstrate sensitivity to other types of legumes. This phenomenon, referred to as allergic cross-reactivity, occurs when a secondary allergen is capable of being recognized as a primary allergen. This recognition triggers the immune system to identify both allergens as either identical or similar, leading to the production of immunoglobulin E (IgE) antibodies or the activation of cellular responses, such as T cell response (Cox, Eigenmann, & Sicherer, 2021; Kamath, Bublin, Kitamura, Matsui, Ito, & Lopata, 2023).

During the production of IF, heat treatments such as pasteurization, sterilization, and spray drying are essential to guarantee microbiological safety and uphold product quality throughout its shelf life. Nevertheless, these processes can lead to structural changes in proteins, such as denaturation and aggregation, as well as glycation in the presence of reducing sugars (Huang, Qu, Hua, Wang, Jia, & Yin, 2023; Phongphisutthinant et al., 2024). These modifications may alter the exposed epitopes, ultimately affecting the immunogenicity of these proteins. For instance, heating can increase allergenicity by inducing protein unfolding to form new epitopes (Pi, Liu, Sun, Ban, Cheng, & Guo, 2023). Another study reported that aggregation-induced irregular molecular shape led to increased IgE-binding sites of β-lactoglobulin, thus

* Corresponding author.
E-mail address: kasper.hettinga@wur.nl (K.A. Hettinga).

https://doi.org/10.1016/j.foodres.2024.114348
Received 12 December 2023; Received in revised form 25 March 2024; Accepted 17 April 2024
Available online 18 April 2024

increasing the allergenicity (Li, Yang, Zou, Shu, Han, & Yang, 2022). On the other hand, heating can also decrease allergenicity through glycation. This occurs through the covalent modification on lysine and arginine residues within epitopes, resulting in a reduced capacity for IgE binding (Bai et al., 2021; Han et al., 2018).

Furthermore, during digestion, proteins undergo enzymatic breakdown into smaller peptides within the gastrointestinal tract. These resulting peptides may have different immunogenicity compared to the intact proteins due to changes in epitope structure (Li, Zhu, Zhou, Peng, & Guo, 2013). Nevertheless, certain peptides derived from digestion, whether from common allergens or other proteins not known to be an allergen, can still retain, or even develop, immunogenic properties. These peptides have the potential to elicit immune responses mediated by B cells and T cells through distinct mechanisms (Cano & Lopera, 2013; Sanchez-Trincado, Gomez-Perosanz, & Reche, 2017). B-cell immune responses involve the recognition of specific peptides by B-cell receptors, leading to the production of antibodies, including IgG, and in the case of an eventual allergic reaction, also IgE (Van Wijk, Hartgring, Koppelman, Pieters, & Knippels, 2004). Once bound to effector cells such as mast cells or basophils, these antibodies can bind to the peptides and trigger allergic reactions upon subsequent exposure. T-cell immune responses, particularly T-helper (Th) cell responses, involve the recognition of exogenous digestion-derived peptides by major histocompatibility complex class II (MHC-II) molecules on the surface of antigen-presenting cells. This recognition activates the differentiation of CD_4^+ T-cells into Th2 cells and initiates the release of cytokines, particularly interleukin-4 (IL-4), which further stimulates B-cell activation and IgE antibody production, as well as up-regulation of MHC-II expression. These immune responses may contribute to the development of allergic reactions and inflammation in individuals with certain protein allergies.

To our knowledge, only a few studies regarded the immunogenic potential of proteins at the peptide level, including e.g., milk proteins (Nutten et al., 2020), rice prolamins (Moreno, Muñoz-Suano, López-Casado, Torres, Sousa, & Cebolla, 2016), peanut (Di Stasio et al., 2017), and soy protein (De Angelis, Pilolli, Bavaro, & Monaci, 2017). However, none of these studies were conducted in the context of infant digestion. Infants possess an immature digestive and immune system, which can make them more susceptible to certain components, such as soy protein, present in soy-based IF. The lowered proteolytic capacity of infants keeps proteins and peptides more intact, and thus the potential immunogenicity may differ from that of adults. Additionally, the impact of different heating processes on peptide immunogenicity has been relatively understudied, warranting further investigation. Therefore, understanding these factors is crucial for evaluating the risk of immune reactions in infants consuming soy-based IF. This can provide insights into the persistence or elimination of immunogenic peptides during infant digestion, thereby facilitating the development of more suitable IF for susceptible individuals.

Therefore, this study aimed to investigate the influence of different heating processes on the potential immunogenicity of soy peptides after static *in vitro* infant gastrointestinal digestion. To achieve this aim, soy protein was heated in wet (in solution) and dry (as powder) conditions and then subjected to one-month *in vitro* infant digestion, followed by peptidomics analysis based on LC-MS/MS. After that, online bioinformatics tools that are time-saving, cost-effective, sensitive, and high-throughput (Wang, Wang, Liu, & Fu, 2020), were used to predict and identify the potential T-cell and linear B-cell epitopic peptides in the digests. The intensities of the epitopic peptides upon different heating and digestion conditions were analyzed compared to the non-treated condition. Finally, all the putative epitopes were further searched in the UniprotKB/Swiss-Prot database to find possible cross-reactive peptides from other legumes.

2. Materials and methods

2.1. Materials and chemicals

Fresh soybean (*Glycine* max; protein content: 39.3 %, w/w) of the Obelix variety was obtained from Wageningen Plant Research, Lelystad, The Netherlands (courtesy ing. Ruud Timmer). All the chemicals were obtained from Sigma-Aldrich (Missouri, USA) unless otherwise stated.

2.2. Preparation of soy protein samples

Soy protein (protein content: 77.9 %, w/w) was extracted, identified, and heat-treated, according to our previously established method (Tang, Wichers, & Hettinga, 2022, 2023). Briefly, soy protein and glucose were initially mixed in powdered form at a weight ratio of 1:4 to mimic the ratio especially found in plant-based IF. Subsequently, for dry heating (D), these powdered mixtures were heated in a desiccator at 60 % humidity and 60 °C for 48 h. The dry-heated samples were then reconstituted in 10 mM PBS buffer (pH 7.4) to achieve a protein concentration of 1.2 % (w/v) to mimic IF. In the case of wet heating (W), the powdered mixtures were first dissolved in PBS buffer to reach a 1.2 % protein content and then heated in a waterbath at 85 °C for 30 min. For the non-treated (NT) samples (controls), mixed powders were directly dissolved in PBS buffer to reach a final 1.2 % protein concentration without any treatment. These heating conditions were selected to obtain substantial structural modifications. All heat treatments were carried out in independent duplicates. Samples were stored at −20 °C until use.

2.3. Structural modification detection

To investigate further structural modifications, differently heated samples, along with the NT samples, were centrifuged (12,000 × g, 20 min, 4 °C) and filtered (0.45 μm RC membrane filter) to remove the insoluble parts. The protein contents of the resulting supernatants were quantified through DUMAS Flash EA 1112 Protein analyzer (Thermo Fisher Scientific, Massachusetts, USA), with a nitrogen-to-protein conversion factor of 5.7. Protein composition was analyzed by SDS-PAGE under both reducing and non-reducing conditions, according to Xiong, Boeren, Vervoort, and Hettinga (2021), with a minor modification involving the loading of 10 μg protein per sample. The formation of advanced glycation end products (AGEs) was measured by intrinsic fluorescence at 370 nm (excitation) and 440 nm (emission), as described by Deng, Govers, Bastiaan-Net, van der Hulst, Hettinga, and Wichers (2019). The results were presented as relative fluorescence of the heated samples compared to the NT sample. The loss of amino groups (%) was quantified by determining blocked amino groups in the heated samples relative to the total free amino groups in the NT sample, based on the o-phthaldialdehyde (OPA) method from Deng et al. (2019). Surface hydrophobicity was measured by the 8-anilino-1-naphthalenesulfonic acid ammonium salt (ANS) fluorescence assay (Tang et al., 2022). The results were displayed as relative fluorescence per heated sample to the NT sample.

2.4. In vitro static infant gastrointestinal digestion

Simulated *in vitro* infant digestion was performed in duplicate based on the protocol of Ménard et al. (2018). All the sample mixtures were submitted to 60 min gastric digestion and 60 min intestinal digestion. Sampling points were 10 min and 60 min after intestinal digestion (I10 and I60). Digestion processes were stopped by immediately transferring into liquid nitrogen. The degree of hydrolysis was measured by an OPA assay according to Mulet-Cabero, Rigby, Brodkorb, and Mackie (2017). All samples that underwent duplicate digestions were analyzed in technical duplicate.

2.5. Peptidomics analysis

To identify the peptides after *in vitro* digestion of wet and dry heated soy protein, as well as those in the NT samples, sample digests were pretreated as described by Dingess et al. (2017) with some modifications. Briefly, spin filters with a 10 kDa cut-off membrane (Sigma-Aldrich, Missouri, USA) were used to fractionate the peptide in the digests, removing intact protein and digestive enzymes. Subsequently, the peptide fractions underwent a purification process through a homemade solid phase extraction (SPE) C18 column, after which they were concentrated to offset any dilution that may have occurred during the SPE stage. The purified peptides were loaded onto a Thermo nLC 1000 system (Thermo, Waltham, MA, USA) coupled to an LTQ orbitrap XL (Thermo Fisher Scientific, Breda, The Netherlands). LC-MS/MS was run by MaxQuant 1.6.3.4 with "unspecific" digestion mode. The instrument parameters were set according to Zenker, Wichers, Tomassen, Boeren, De Jong, and Hettinga (2020).

A soy database (UP000008827) containing 74,863 protein sequences downloaded from Uniprot (https://www.uniprot.org, accessed June 2022) was used to analyze in MaxQuant, together with another database for common contaminants, including e.g., bovine trypsin (P00760), porcine trypsin (P00761), and human Keratin K2C1 (P04264). Peptides were identified with unspecific enzyme cleavage and with lengths from 8 to 25 AAs. To conduct the analysis, peptides that were present in both duplicate digestions were considered valid. Peptides with a score ≥ 80 were included in the overall profiles, while only those peptides with a score ≥ 100 were considered for epitopic analysis.

2.6. T-cell epitope prediction

T-cell epitopes in the digests were predicted by using the online IEDB MHC-II Binding Prediction tool (https://tools.iedb.org/mhcii/, accessed April 2023), which specifically targets the exogenous antigens, e.g., digestion-derived peptides. MHC-II binding epitopes were reported to range from 15 to 24 AAs in length (Knol, de Jong, Ulfman, & Tiemessen, 2019). Peptides falling within these length range, as identified by LC-MS/MS, were submitted to the prediction software. The default "IEDB recommended 2.22" method was chosen, and 27 HLA class II alleles were used for the prediction. Two key factors were used to filter the results: percentile ranking value and inhibitory concentration 50 (IC50) value. Both values represented the binding affinity between peptides and alleles, with lower values indicating higher binding affinity. Therefore, to screen the potential T-cell epitopes, the maximum percentile ranking value was set to 10, and the maximum IC50 score was set to 250 nM, according to Zenker et al. (2020) and Zhou, He, Zhang, Wang, Sun, and Liu (2022) with some modifications.

Subsequently, these potential epitopes were applied to an online IL4pred tool (https://webs.iiitd.edu.in/raghava/il4pred/predict.php, accessed April 2023,) to predict the ability to induce IL-4 secretion by Th2 cells (Dhanda, Gupta, Vir, & Raghava, 2013). Prediction mode was set as hybrid (SVM + motif), and an SVM threshold of 0.2 was selected.

2.7. Linear B-cell epitope identification

A soy B-cell epitope database was downloaded (https://www.iedb.org/, accessed March 2023) from the IEDB database (Vita et al., 2018). The search settings were: "linear peptide", with "human" specified as the host, and "*Glycine* max" as the organism. The search yielded a total of 550 epitopes. Notably, B-cell epitopes are known to exist in both linear and conformational forms. However, due to the data available, which consisted of peptide sequences formed during infant digestion, we focused on linear epitopes specifically. The peptide sequences in the digests were compared against this database, and only those peptides with a length longer than or equal to that of soy B-cell epitopes were reported. Moreover, a further filter was applied to select peptides with a matching rate ≥ 60 %. The matching rate refers to the ratio of the length of B-cell epitopes to that of the peptides in the digests. In general, a higher matching rate indicates a higher likelihood that the peptides may serve as B-cell epitopes and could be recognized by the receptors.

2.8. Alignment of potential epitopic peptides with other legume proteins

Both potential epitopic peptides of T-cell and B-cell were applied to the online BLASTp tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE = Proteins, accessed April 2023) to search for similar peptide sequences in other legume species. The search was conducted in the UniprotKB/Swiss-Prot database. The maximum number of aligned sequences was set to 50, and the expect threshold was set to be 0.05. The results were filtered to include only sequences containing at least six consecutive and no more than three different AAs, as well as to exclude any peptides with a percent identity score lower than 80. In most situations, the cross-reactivity requires a percent identity score over 70 (Aalberse, 2000; Bublin & Breiteneder, 2020), but to increase reliability, in our study, this number was set to 80. Percent identity is a quantitative result, as defined by the percentage of the number of identical AAs in matching peptides from other legumes to the length of the aligned sequence in the soy epitopic peptides.

2.9. Statistical analysis

All the data were visualized by GraphPad Prism 9.0 and analyzed based on R 4.2.1 and RStudio 2022.07.0. Statistical differences were analyzed using ANOVA through IBM SPSS 25.0 and GraphPad Prism 9.0, with $p < 0.05$ being considered significant.

3. Results and discussion

3.1. Effect of different heating on structural modification and infant digestion

Structural modifications of the wet-heated (W) and dry-heated (D) samples were evaluated in comparison to the non-treated (NT) sample, including AGE formation, loss of amino groups (%), protein composition, and surface hydrophobicity (Fig. 1). The level of glycation was assessed by the AGE formation (Fig. 1A) and the loss of amino groups (%) (Fig. 1B), indicating that only the D sample was significantly glycated ($p < 0.01$). SDS-PAGE results (Fig. 1C) showed that in the W sample, all subunits remained unchanged under the reducing condition. Furthermore, under the non-reducing condition, disulfide rearrangements were observed, as evidenced by the decreased intensity of the disulfide-linked glycinin subunits and the formation of soluble aggregates in the wells at the top of the non-reducing gel. Notably, only the D sample exhibited a considerable shifting up and disappearance of the subunits (Fig. 1C) due to intense glycation between soy protein and glucose (Fig. 1 A&B). In addition, Fig. 1D revealed that the surface hydrophobicity of soy protein was significantly increased after wet heating ($p < 0.001$), while it was decreased after dry heating ($p < 0.05$). According to our earlier studies, this is because wet heating induced protein unfolding and thereby increased the surface hydrophobicity, while dry heating led to glycation and the formation of hydrophilic AGEs on the protein surface and thus resulted in decreased surface hydrophobicity (Tang et al., 2022, 2023).

The degree of hydrolysis (DH) after *in vitro* digestion is shown in Fig. 2. At 10 min of intestinal digestion (I10), the DH of the NT, W, and D samples were 38.7 %, 45.4 %, and 33.0 %, respectively. As digestion progressed to 60 min (I60), their DH values were increased to 42.7 %, 55.0 %, and 40.7 %, respectively. These findings demonstrate an increase in the DH during prolonged digestion, as well as highlighting the enhancing effect of wet heating on infant digestion and the inhibitory effect of dry heating. Nonetheless, no significant difference was found in the DH between the NT and D samples at I60. These results corroborate the conclusions drawn from our previous research (Tang et al., 2022,

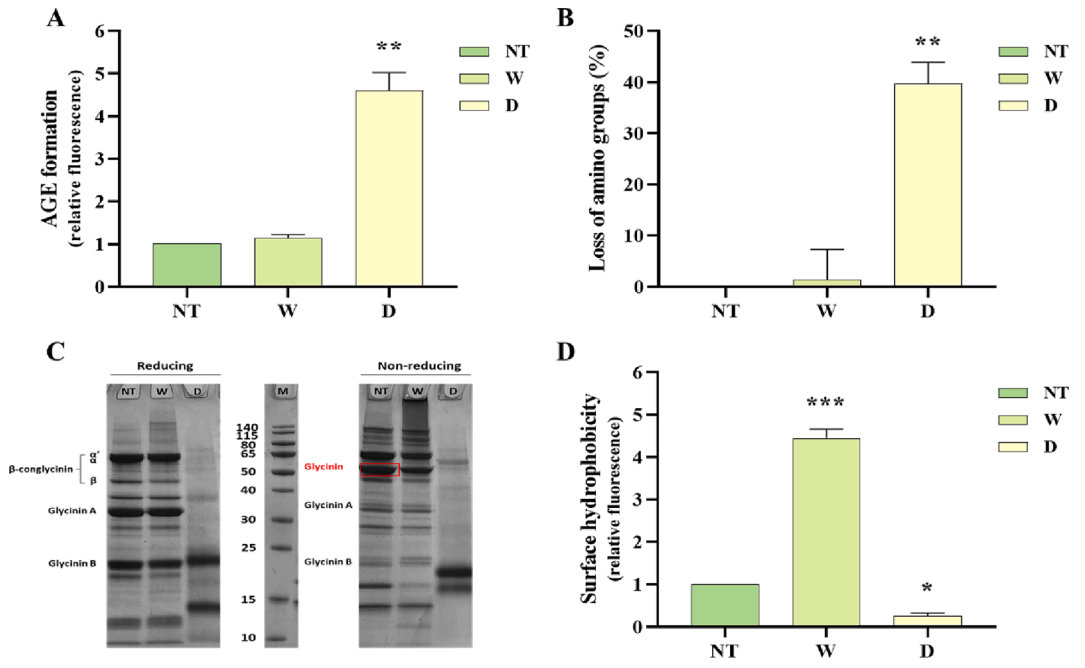

Fig. 1. Structural modifications of soy protein under wet (W) or dry (D) heating, with non-treated (NT) samples as control. (A) AGE formation, (B) loss of amino groups (%), (C) protein composition, with disulfide-linked glycinin subunits in red, and (D) surface hydrophobicity. Statistical differences were analyzed with ANOVA and Dunnett's test: *p < 0.05; **p < 0.01; ***p < 0.001.

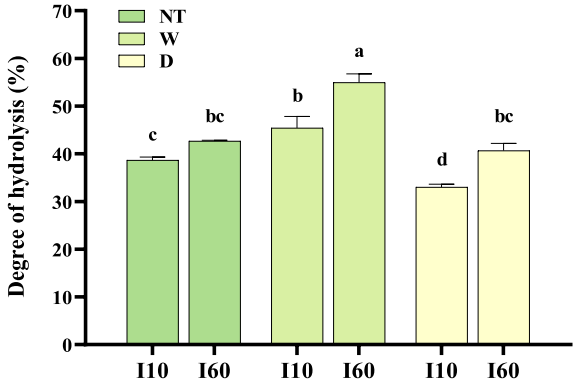

Fig. 2. Degree of hydrolysis (%) of soy protein under wet (W) or dry (D) heating, after *in vitro* infant digestion, with non-treated (NT) samples as control. I10 and I60 represent 10 min and 60 min of intestinal digestion, respectively. Statistical differences were analyzed with ANOVA and Duncan post-hoc test. Different letters above the bars represent significant differences (p < 0.05).

2023). Specifically, wet heating promoted protein unfolding and exposed enzyme-cleavage sites to enhance digestion. In contrast, despite that glycation-induced aggregation upon dry heating led to the burial of these sites, the presence of multiple enzymes in pancreatin negated the detrimental effects of glycation, as evidenced by the similar DH between the NT and D samples observed at I60.

3.2. Peptide identification in the digests

Peptide release after 10 min and 60 min of intestinal digestion upon different heating conditions was analyzed by LC-MS/MS. A total of 8,546 peptides were detected, although only 6,684 peptides with a score over 80 were retained for further analysis. Details of peptide intensity and number were demonstrated in Fig. 3. Due to the limitation of the analysis method, which only targets peptides with 8 to 25 amino acids (AAs), longer or shorter peptides were not detected. Therefore, the results do not directly reflect the degree of digestion.

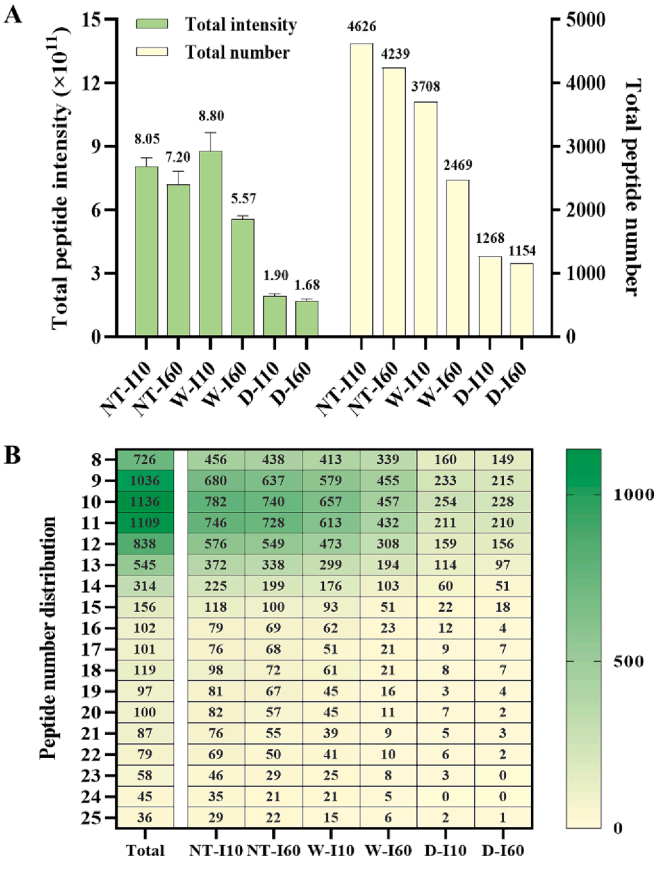

Fig. 3. (A) Total intensity and number of peptides. (B) Peptide number distribution with length from 8 to 25 amino acids. Total refers to all the diverse peptides found in all the samples. Peptides were fractionated from the digests under wet (W) or dry (D) heating, with non-treated (NT) samples as control. I10 and I60 represent 10 min and 60 min of intestinal digestion, respectively.

Fig. 3A illustrates the total intensity and number of peptides upon different heating and digestion conditions. Total intensity represents the overall abundance or signal strength of peptides in a given heating and digestion condition, with the higher total intensity signifying a higher cumulative abundance of peptides. Total peptide number represents the diversity of peptides detected, and each identified peptide contributes to the count, regardless of its abundance. In general, both the total intensity and number declined from NT to W to D, as well as from I10 to I60. Specifically, the total intensity (Fig. 3A, left), showed that the D sample had the lowest intensity among all the samples, being 4.2 and 4.3 times lower than that of the NT sample at I10 and I60, respectively. In addition, with prolonged digestion, the NT and D samples experienced a reduction of 10 % and 12 % in the total intensity, respectively, whereas the W sample exhibited the highest decrease of 37 %. Regarding the total peptide number (Fig. 3A, right), the NT sample exhibited the highest peptide count, followed by the W sample, while the D sample had the lowest amount. Their reductions from I10 to I60 were 8 %, 33 %, and 9 % for the NT, W, and D samples, respectively, showing the most substantial decrease in the W sample, similar to the intensity.

In general, the identifiability of the peptides decreased after wet heating and reached the lowest level after dry heating, compared to the NT condition. For the W sample, this decrease may be explained by improved digestion through protein unfolding, resulting in shorter peptides with lengths less than 8 AAs that were undetectable by the LC-MS/MS method used. On the other hand, the lowest identifiability of the peptides in the D sample may be due to the potential formation of cross-linked compounds like pentosidine on the protein surface through glycation. These compounds may persist in the peptide fragments even after digestion, hindering their detection by LC-MS/MS. Moreover, with extended digestion, the total peptide intensity and count also decreased due to the further cleavage of larger peptides into smaller ones.

Peptides might be potential antigens to trigger immune responses mediated by B cells and T cells. Peptide length is important as it affects the accessibility and recognition of B-cell epitopes as well as the binding and activation of T-cell receptors. B-cell epitopes are mainly 5–22 AAs long (Singh, Ansari, & Raghava, 2013), whereas MHC-II binding T-cell epitopes are usually ranging from 15 to 24 AAs (Knol et al., 2019). Fig. 3B illustrates the peptide number distribution in the length range of 8–25 AAs, providing information on the peptide diversity that may be responsible for the potential immunogenicity after digestion. It was observed that approximately 80 % of the total peptides were 8–13 AAs long. This percentage varied between 78 % and 91 % among different heating and digestion conditions. In addition, around 98 % of the peptides were in the length range of B-cell epitopes (in our study, 8–22 AAs); while only 14 % matched the length criteria for T-cell epitopes (15–24 AAs). This indicates that the immune response mediated by B cells may be more readily activated by digestion-derived peptides, than a response mediated by T cells. Nevertheless, apart from peptide length, peptide sequence is crucial for the recognition and activation of immune cells as well. This is because specific AA sequence of the peptide can determine its binding affinity to the T-cell receptors or B-cell receptors. Hence, all peptides that meet the length criteria for epitopes were further studied using a sequence analysis, which will be described in sections 3.3 and 3.4.

3.3. T-cell epitope prediction results

As shown in Table 1, a total of 29 peptides in the digests were predicted to be potential MHC-II binding T-cell epitopes. Among these peptides, a substantial portion originated from specific proteins, with lipoxygenase accounting for 7 peptides, followed by alcohol dehydrogenase, β-conglycinin, and glycinin, each contributing 4 peptides. Conversely, Bet v I/Major latex protein domain-containing protein, oleosin, and other proteins yielded only 1 peptide each. Amongst all, glycinin (Gly m 6), β-conglycinin (Gly m 5), and Bet v I/Major latex protein domain-containing protein (Gly m 4) are considered the main soy allergens. Therefore, despite that common soy allergens were considerably hydrolyzed after intestinal digestion, intact linear epitopic sequences were still present in the digests, which may potentially elicit immune reactions. Interestingly, lipoxygenase, alcohol dehydrogenase, and other proteins, which are not seen as common soy allergens, exhibited potential immunogenicity after digestion as well. This intriguing finding may be attributed to the broader concept of immunogenicity beyond just allergenicity. While certain food proteins may not cause allergic responses, they might still trigger immune system activation, antibody production, or other immunological events (Leviatan, Vogl, Klompus, Kalka, Weinberger, & Segal, 2022; Van Wijk et al., 2004).

Moreover, regarding the length of the potential T-cell epitopes, 10 peptides consisted of 15 AAs, while 16 peptides were 16–19 AAs long. These specific lengths accounted for almost 90 % of the total 29 peptides. Thus, despite that the length criterium for MHC-II binding T-cell epitopes was 15–24 AAs, peptides in the soy digests with 15–19 AAs were more often predicted to be T-cell epitopes. This might be due to the low number of longer peptides present in the digests. Furthermore, the peptide with sequence "LSIVDMNEGALLLPH" (P0DO15, AA 440–454) from β-conglycinin can bind to 4 different HLA II alleles, indicating that a larger population may have the potential to recognize and interact with this particular peptide.

Afterward, the IL4 pred tool was used for further analysis, and only the peptides with prediction scores over 0.2 were seen as potential IL-4 inducers. A higher prediction score indicates a greater predicted ability to stimulate Th2 cells to secrete IL-4. Results show that 13 out of 29 peptides were predicted to be IL-4 inducers. Amongst these peptides, the peptide "DLEGKTVGTVGAGRIGK" (I1N5S0, AA 199–215) had the highest score, followed by "IDTNSFQNQLDQMPR" (P11828, AA 163–177), "IKQVHQLNANPDVHG" (C6TMV9, AA 79–93), and "KEIFRTDGEQALKFPPPK" (B3TDK6, AA 332–349). Moreover, peptides "GHEASGIVESVGKGV" (A0A0R4J4U4, AA 68–82) and "PRIFGHEAGGIVESVGEGV" (I1KAJ4, AA 66–84) could bind to three different HLA II alleles, suggesting that they could cause more people to be allergic than other epitopes.

Overall, 29 potential T-cell epitopes were found in the digests, with 13 of them recognized as IL-4 inducers. Generally, the epitopic peptides were more likely to have 15–19 AAs. In addition, aside from the 9 peptides derived from the main soy allergens, an additional 20 peptides from minor or non-allergic soy proteins demonstrated immunogenic potential as well.

3.4. B-cell epitope prediction results

By aligning the peptide sequences to the soy linear B-cell epitopes database from IEDB, 116 peptides were found, with 27 peptides with a matching rate over 60 % being filtered as potential B-cell epitopes (Table 2). Among these, 10 peptides were derived from glycinin (Gly m 6), 10 peptides from 2S albumin (Gly m 8), and 1 peptide from β-conglycinin (Gly m 5), which are known as common soy allergens. Furthermore, 3 peptides from lectin and 3 peptides from basic 7S globulin were also recognized as B-cell epitopes, indicating that some minor soy allergens still had immunogenic potential after *in vitro* digestion.

In terms of the length distribution of linear B-cell epitopes, 21 out of 27 peptides had 8–11 AAs. It suggests that the shorter peptides were relatively easier to be recognized by B cells, while longer peptides may contain multiple epitopes or have complex structures that hinder binding to B-cell receptors. Furthermore, it is noteworthy that only two peptides demonstrated a perfect matching rate of 100 %, showing that these peptides have already been identified as B-cell epitopes before. One peptide is "FSFREQPQ" from glycinin (P11828, AA 20–27) with the epitope ID 913228, and another peptide is "NQSEELEEK" from albumin (P19594, AA 120–128) that matches the epitope ID: 913848.

To sum up, 27 out of 116 peptides were seen as potential linear B-cell

Table 1
List of potential T-cell epitopes recognized in the digests, along with HLA class II alleles and other relevant information.

#	Potential T-cell Epitope	Protein Group	Protein ID	Position	Length	HLA class II alleles	Method	IL4 Pred	IL4 Inducer
1	DLEGKTVGTVGAGRIGK	Formate dehydrogenase, mitochondrial	I1N5S0	199–215	17	HLA-DQA1*05:01/ HLA-DQB1*03:01	Consensus (comb. lib./smm/nn)	1.416	Yes
2	IDTNSFQNQLDQMPR	Glycinin G3	P11828	163–177	15	HLA-DRB1*04:05	Consensus (smm/ nn/sturniolo)	1.277	Yes
3	IKQVHQLNANPDVHG	Methenyltetrahydrofolate cyclohydrolase	C6TMV9	79–93	15	HLA-DRB1*13:02	Consensus (smm/ nn/sturniolo)	1.266	Yes
4	KEIFRTDGEQALKFPPPK	Lipoxygenase	B3TDK6	332–349	18	HLA-DRB3*01:01	Consensus (comb. lib./smm/nn)	1.245	Yes
5	KEIFRTDGEQALKFPPPKVIQ	Lipoxygenase	B3TDK6	332–352	21	HLA-DRB3*01:01	Consensus (comb. lib./smm/nn)	1.225	Yes
6	GHEASGIVESVGKGV	Alcohol dehydrogenase 1	A0A0R4J4U4	68–82	15	HLA-DQA1*05:01/ HLA-DQB1*03:01/ HLA-DRB1*13:02	Consensus (comb. lib./smm/nn), Consensus (smm/ nn/sturniolo)	1.037	Yes
7	PRIFGHEAGGIVESVGEGV	Alcohol dehydrogenase	I1KAJ4	66–84	19	HLA-DQA1*05:01/ HLA-DQB1*03:01/ HLA-DRB1*13:02	Consensus (comb. lib./smm/nn), Consensus (smm/ nn/sturniolo)	0.967	Yes
8	NSIPAFRKGSIPGVTSDHM	Uncharacterized protein	C6T0B5	46–64	19	HLA-DQA1*05:01/ HLA-DQB1*03:01	Consensus (comb. lib./smm/nn)	0.308	Yes
9	MRKNVLDFNSVADLT	Lipoxygenase	I1M596	26–40	15	HLA-DRB1*13:02	Consensus (smm/ nn/sturniolo)	0.305	Yes
10	SKHFLAQSFNTNEDTAE	Cupin type-1 domain-containing protein	A0A0R0GMV1	233–249	17	HLA-DPA1*01:03/ HLA-DPB1*04:01	NetMHCIIpan	0.300	Yes
11	SKHFLAQSFNTNEDIAE	Glycinin G4	P02858	234–250	17	HLA-DPA1*01:03/ HLA-DPB1*04:01	NetMHCIIpan	0.288	Yes
12	SVDINEGALLLPHFNSKA	Beta-conglycinin beta subunit 2	F7J077	276–293	18	HLA-DRB1*13:02	Consensus (smm/ nn/sturniolo)	0.253	Yes
13	SSVDINEGALLLPHFNSKA	Beta-conglycinin beta subunit 2	F7J077	275–293	19	HLA-DRB1*13:02	Consensus (smm/ nn/sturniolo)	0.238	Yes
14	SVENVEGNGGPGTIK	Bet v I/Major latex protein domain-containing protein	C6T588	40–54	15	HLA-DQA1*05:01/ HLA-DQB1*03:01	Consensus (comb. lib./smm/nn)	0.172	No
15	GHEAGGIVESVGEGV	Alcohol dehydrogenase	I1KAJ4	70–84	15	HLA-DQA1*05:01/ HLA-DQB1*03:01/ HLA-DRB1*13:02	Consensus (comb. lib./smm/nn), Consensus (smm/ nn/sturniolo)	0.018	No
16	GHEASGIVESVGEGV	Alcohol dehydrogenase 1	I1MA91	60–74	15	HLA-DQA1*05:01/ HLA-DQB1*03:01/ HLA-DRB1*13:02	Consensus (comb. lib./smm/nn), Consensus (smm/ nn/sturniolo)	−0.001	No
17	YREEELKSLRGNGTGE	Lipoxygenase	A0A0R0H569	161–176	16	HLA-DRB1*01:01	Consensus (comb. lib./smm/nn)	−0.019	No
18	VEYREEELKSLRGNGTGE	Lipoxygenase	A0A0R0H569	159–176	18	HLA-DRB1*01:01	Consensus (comb. lib./smm/nn)	−0.058	No
19	NSLLNALPEEVIQHT	Glycinin G1	P04776	449–463	15	HLA-DRB1*09:01	Consensus (comb. lib./smm/nn)	−0.068	No
20	KTNDTPMIGTLAGANS	Glycinin G1	P04776	435–450	16	HLA-DQA1*05:01/ HLA-DQB1*03:01	Consensus (comb. lib./smm/nn)	−0.092	No
21	VEYREEELKSLRGNGTGERKEYD	Lipoxygenase	A0A0R0H569	159–181	23	HLA-DRB1*01:01	Consensus (comb. lib./smm/nn)	−0.118	No
22	YREEELKSLRGNGTGERKEYD	Lipoxygenase	A0A0R0H569	161–181	21	HLA-DRB1*01:01	Consensus (comb. lib./smm/nn)	−0.190	No
23	AFPGSAQAVEKLLKNQ	Beta-conglycinin alpha subunit 2	P0DO15	558–573	16	HLA-DQA1*05:01/	Consensus (comb. lib./smm/nn)	−0.237	No

(*continued on next page*)

Table 1 (*continued*)

#	Potential T-cell Epitope	Protein Group	Protein ID	Position	Length	HLA class II alleles	Method	IL4 Pred	IL4 Inducer
24	LADELINAAKGSSNS	Ribosomal protein S7 domain-containing protein	I1L2H4	173–187	15	HLA-DQB1*03:01 HLA-DQA1*05:01/ HLA-DQB1*03:01	Consensus (comb. lib./smm/nn)	−0.278	No
25	SDVLTDAKEKLPADKA	SMP domain-containing protein	I1L849	201–216	16	HLA-DRB1*03:01	Consensus (smm/ nn/sturniolo)	−0.325	No
26	LSIVDMNEGALLLPH	Beta-conglycinin alpha subunit 2	P0DO15	440–454	15	HLA-DQA1*01:02/ HLA-DQB1*06:02/ HLA-DRB1*13:02/ HLA-DRB3*02:02	Consensus (comb. lib./smm/nn), Consensus (smm/ nn/sturniolo), NetMHCIIpan	−0.791	No
27	THRYEAGVVPPGARFEPPR	Oleosin	I1N747	16–34	19	HLA-DQA1*05:01/ HLA-DQB1*03:01	Consensus (comb. lib./smm/nn)	−0.903	No
28	NGGVVLGADSRTSTGV	Proteasome subunit beta	I1K964	23–38	16	HLA-DQA1*05:01/ HLA-DQB1*03:01	Consensus (comb. lib./smm/nn)	−1.115	No
29	KIAQGVVGAVKDKGSV	Inosine-5′-monophosphate dehydrogenase	I1MJD7	426–441	16	HLA-DQA1*05:01/ HLA-DQB1*03:01	Consensus (comb. lib./smm/nn)	−1.151	No

Table 2
Comparison of peptides identified in the digests and B-cell epitopes reported in the IEDB database, with a matching rate over 60% being included in the list. The peptide sequence in bold matches 100% with the B-cell epitopes reported.

#	Peptide	Peptide length	Matching B-cell Epitope	Matching rate (%)	Epitope ID	Protein Group	Protein ID	Epitope Position
1	**FSFREQPQ**	8	FSFREQPQ	100	913228	Glycinin G3	P11828	20–27
2	**NQSEELEEK**	9	NQSEELEEK	100	913848	2S Albumin	P19594	120–128
3	ENQSEELEEK	10	NQSEELEEK	90.0	913848	2S Albumin	P19594	120–128
4	NQSEELEEKQ	10	NQSEELEEK	90.0	913848	2S Albumin	P19594	120–128
5	IYIQQGSGI	9	YIQQGSGI	88.9	914294	Glycinin G3	P11828	92–99
6	SFASNLPH	8	FASNLPH	87.5	913189	Lectin	P05046	260–266
7	MENQSEELEEK	11	NQSEELEEK	81.8	913848	2S Albumin	P19594	120–128
8	GGLSVISPK	9	GLSVISP	77.8	913280	Glycinin G4	P02858	268–274
9	SFASNLPHA	9	FASNLPH	77.8	913189	Lectin	P05046	260–266
10	IMENQSEELEEK	12	NQSEELEEK	75.0	913848	2S Albumin	P19594	120–128
11	LGTRQLEE	8	GTRQLE	75.0	913313	Basic 7S globulin 2	Q8RVH5	392–397
12	QQFHNQDI	8	QQFHNQ	75.0	913988	Basic 7S globulin 2	Q8RVH5	237–242
13	FSFREQPQQNE	11	FSFREQPQ	72.7	913228	Glycinin G3	P11828	20–27
14	LFGEEEEQRQ	10	LFGEEEE	70.0	913609	Beta-conglycinin beta subunit 1	P25974	196–202
15	EGGLSVISPK	10	GLSVISP	70.0	913280	Glycinin G4	P02858	268–274
16	SFASNLPHAS	10	FASNLPH	70.0	913189	Lectin	P05046	260–266
17	KIMENQSEELEEK	13	NQSEELEEK	69.2	913848	2S Albumin	P19594	120–128
18	MENQSEELEEKQK	13	NQSEELEEK	69.2	913848	2S Albumin	P19594	120–128
19	IMENQSEELEEKQK	14	NQSEELEEK	64.3	913848	2S Albumin	P19594	120–128
20	MENQSEELEEKQKK	14	NQSEELEEK	64.3	913848	2S Albumin	P19594	120–128
21	SFQNQLDQMPR	11	FQNQLDQ	63.6	913224	Glycinin G3	P11828	168–174
22	VEGGLSVISPK	11	GLSVISP	63.6	913280	Glycinin G4	P02858	268–274
23	YQGNSGPLVNP	11	NSGPLVN	63.6	913858	Glycinin G5	P04347	510–516
24	APEFLEHA	8	FLEHA	62.5	913209	Glycinin G3	P11828	228–232
25	NALEPDHR	8	LEPDH	62.5	913606	Glycinin G4	P02858	40–44
26	KIMENQSEELEEKQK	15	NQSEELEEK	60.0	913848	2S Albumin	P19594	120–128
27	VTLGTRQLEE	10	GTRQLE	60.0	913313	Basic 7S globulin 2	Q8RVH5	392–397

epitopes, with 2 of them already documented in the database. Out of these, 21 peptides were derived from common soy allergens, while 6 peptides originated from less allergic soy proteins. The majority of these potential epitopes had lengths ranging from 8 to 11 AAs.

3.5. Evaluation of potential cross-reactivity compared to other legumes

To identify cross-reactivity between soy proteins and other legume proteins, the potential epitopic peptides were aligned to the UniprotKB/Swiss-Prot database via a BLASTp search. By IEDB bioinformatics tools, we have already predicted potential T-cell epitopic peptides in the digests, as well as identified B-cell epitope-containing peptides from the database (sections 3.3 and 3.4). But to increase the reliability, we only selected 13 potential T-cell epitopic peptides that were IL-4 inducers and 2 linear B-cell epitopes that had a 100 % match.

The results in Table 3 show that several T-cell epitopic peptides shared identical or similar sequences with other legumes, while no B-cell epitopic peptides did so. In total, 28 matching peptides from other

Table 3
Comparison of potential epitopes in the digests and proteins belonging to other legume species. All the matching peptides are obtained through BLASTp search in the UniprotKB/Swiss-Prot database, with at least six consecutive and no more than three different amino acids compared to epitopic peptides. The peptide sequence in bold matches 100% with the epitopic peptide. Different amino acids are highlighted in red.

Epitopic Peptide	Protein Group	Matching Peptide	Identity	Protein Group	Protein ID	Position	Organism	General Name
IDTNSFQNQLDQMPR	Glycinin G3	IDTSSFQNQLDQMPR	93.3	Legumin	Q9SMJ4	165–179	*Cicer arietinum*	Chickpea
KEIFRTDGEQALKFPPPK	Lipoxygenase	KEIFRTDGENTLKYPPPK	83.3	Seed linoleate 9S-lipoxygenase-3	P09918	336–353	*Pisum sativum*	Garden pea
		KEIFRTDGEQVLKFTPP	88.2	Seed linoleate 9S-lipoxygenase-2	P14856	340–356	*Pisum sativum*	Garden pea
		KEILRTDGEQVLKFPPP	88.2	Linoleate 9S-lipoxygenase 1	P27480	337–353	*Phaseolus vulgaris*	Kidney bean
		ELFRSDGEAALKFPPPK	82.4	Linoleate 9S-lipoxygenase	P38414	341–357	*Lens culinaris*	Lentil
KEIFRTDGEQALKFPPPKVIQ	Lipoxygenase	KEIFRTDGENTLKYPPPKVIQ	85.7	Seed linoleate 9S-lipoxygenase-3	P09918	336–356	*Pisum sativum*	Garden pea
		KEIFRTDGEQVLKFTPPHVI	85.0	Seed linoleate 9S-lipoxygenase-2	P14856	340–359	*Pisum sativum*	Garden pea
		KEILRTDGEQVLKFPPPHVI	85.0	Linoleate 9S-lipoxygenase 1	P27480	337–356	*Phaseolus vulgaris*	Kidney bean
		ELFRSDGEAALKFPPPKVIQ	85.0	Linoleate 9S-lipoxygenase	P38414	341–360	*Lens culinaris*	Lentil
GHEASGIVESVGKGV	Alcohol dehydrogenase 1	GHEAGGIVESVGEGV	86.7	Alcohol dehydrogenase 1	P12886	69–83	*Pisum sativum*	Garden pea
		GHEAAGIVESVGEGV	86.7	Alcohol dehydrogenase class-3	P80572	67–81	*Pisum sativum*	Garden pea
PRIFGHEAGGIVESVGEGV	Alcohol dehydrogenase	**PRIFGHEAGGIVESVGEGV**	100	Alcohol dehydrogenase 1	P12886	65–83	*Pisum sativum*	Garden pea
		ILGHEAAGIVESVGEGV	88.2	Alcohol dehydrogenase class-3	P80572	65–81	*Pisum sativum*	Garden pea
SSVDINEGALLLPHFNSKA	Beta-conglycinin beta subunit 2	INEGALLLPHYNSKA	93.3	Conglutin beta 2	Q6EBC1	371–385	*Lupinus albus*	White lupine
		EINEGALLLPHYNSKA	87.5	Conglutin beta 1	Q53HY0	370–385	*Lupinus albus*	White lupine
		EINEGALLLPHYNSKA	87.5	Conglutin beta 1	F5B8V9	441–456	*Lupinus angustifolius*	Narrow-leaved blue lupine
		EINEGALLLPHYNSKA	87.5	Conglutin beta 7	F5B8W5	439–454	*Lupinus angustifolius*	Narrow-leaved blue lupine
		EISEGALLLPHYNSKA	81.3	Conglutin beta 5	F5B8W3	472–487	*Lupinus angustifolius*	Narrow-leaved blue lupine
		IEINEGALLLPHYNSKA	82.4	Conglutin beta 6	F5B8W4	431–447	*Lupinus angustifolius*	Narrow-leaved blue lupine
		IEINEGALLLPHYNSKA	82.4	Conglutin beta 4	F5B8W2	428–444	*Lupinus angustifolius*	Narrow-leaved blue lupine
		IEINEGALLLPHYNSKA	82.4	Conglutin beta 3	F5B8W1	418–434	*Lupinus angustifolius*	Narrow-leaved blue lupine
		VEIKEGALMLPHFNSKA	82.4	Allergen Ara h 1	P43238	436–452	*Arachis hypogaea*	Peanut
		VEIKEGALMLPHFNSKA	82.4	Allergen Ara h 1	P43237	428–444	*Arachis hypogaea*	Peanut
SVDINEGALLLPHFNSKA	Beta-conglycinin beta subunit 2	EINEGALLLPHYNSKA	87.5	Conglutin beta 1	F5B8V9	441–456	*Lupinus angustifolius*	Narrow-leaved blue lupine
		EINEGALLLPHYNSKA	87.5	Conglutin beta 7	F5B8W5	439–454	*Lupinus angustifolius*	Narrow-leaved blue lupine
		IEINEGALLLPHYNSKA	82.4	Conglutin beta 6	F5B8W4	431–447	*Lupinus angustifolius*	Narrow-leaved blue lupine
		IEINEGALLLPHYNSKA	82.4	Conglutin beta 4	F5B8W2	428–444	*Lupinus angustifolius*	Narrow-leaved blue lupine
		IEINEGALLLPHYNSKA	82.4	Conglutin beta 3	F5B8W1	418–434	*Lupinus angustifolius*	Narrow-leaved blue lupine

legume proteins were identified to align to 7 potential T-cell epitopic peptides. Such alignments, in order of frequency, were observed in *Lupinus angustifolius* (11 peptides, narrow-leaved blue lupine), *Pisum sativum* (8, pea), *Lens culinaris* (2, lentil), *Phaseolus vulgaris* (2, kidney bean), *Lupinus albus* (2, white lupine), *Arachis hypogaea* (2, peanut), and *Cicer arietinum* (1, chickpea). Interestingly, if an epitope is from one specific protein, its matching peptides were always derived from a protein with the same or similar sequence and function. For instance, soy-derived peptides from lipoxygenase were aligned to that of *Pisum sativum*, *Phaseolus vulgaris*, and *Lens culinaris*, while peptides from soy β-conglycinin matched to conglutin from *Lupinus* family and matched to Ara h 1 from *Arachis hypogaea*. In addition, of all the identified peptides, the sequence "PRIFGHEAGGIVESVGEGV" from alcohol dehydrogenase was found to be identical in both soy and pea. Thus, this peptide is more likely to cause cross-reactivity.

Many studies have demonstrated that cross-reactivity can occur between soy and other legumes (or legume-based food additives), such as peanut, pea, chickpea, lima bean, green bean, white bean, lentil, guar gum, carob bean, tragacanth, and liquorice (L'Hocine & Boye, 2007; Verma, Kumar, Das, & Dwivedi, 2013). According to our results, apart from peanut, pea, chickpea, and lentil that were already mentioned above, potential epitopic peptides from kidney bean and lupine family may cause cross-reactions as well. However, these results are based on prediction *in silico*, thus further confirmation is required, e.g., through serum testing or similar experiments. Despite so, these results indicate that consumption of other legumes may result in the generation of cross-reactive peptides upon digestion, potentially triggering allergic reactions even in individuals who have only been sensitized to soy allergens.

3.6. Immunogenic potential after digestion upon different heating

To evaluate the immunogenic potential after *in vitro* infant digestion under different heating conditions, we specifically selected 15 potential epitopic peptides as described in section 3.5. The alterations in the intensity of these epitopic peptides are illustrated in Fig. 4. Please note that the intensity of a peptide reflects its cumulative abundance, which can be influenced by factors such as peptide length and its specific sequence. Therefore, it is important to only compare the intensities of the same peptide across different experimental conditions, but not make direct comparisons of intensity between different peptides as the factors affecting their abundance may vary.

For potential T-cell epitopes, in general, wet, but especially dry heating decreased their intensities. Based on that, digestion would lead to a further decrease, indicating a reduction in the likelihood of triggering immunogenic responses mediated by T cells. To be specific, without heating, all 13 peptides had detectable intensities at I10. After wet heating, 12 peptides experienced a decrease in intensity, with 7 of them becoming non-detectable. Similarly, after dry heating, the intensity of 12 out of 13 peptides became non-detectable. Notably, the peptide "IDTNSFQNQLDQMPR" had contrasting trends between wet and dry heating. Its intensity significantly rose by 32 % after wet heating but decreased by around 40-fold after dry heating, compared to the non-treated condition. In addition, the peptide "DLEGKTVGTVGAGRIGK" demonstrated the most significant decline, with its intensity reducing from 2.6×10^8 at I10 to non-detectable at I60 after both wet and dry heating. As digestion progressed from I10 to I60, the intensity of 12 peptides in the NT sample showed a decrease, with 8 of them eventually becoming non-detectable. However, 1 peptide, "GHEASGIVESVGKGV", showed an opposite trend, with an increase of 4.9-fold in the intensity from I10 to I60. In the case of the heated samples at I10, the intensities of most peptides were already decreased due to the heating process, except for the peptide "IDTNSFQNQLDQMPR" mentioned earlier. Only 6 peptides in the wet-heated sample and 1 peptide in the dry-heated sample had a detectable intensity at I10. However, after prolonged digestion, all of these peptides became non-detectable at I60, regardless of whether heating had initially increased or decreased their intensity.

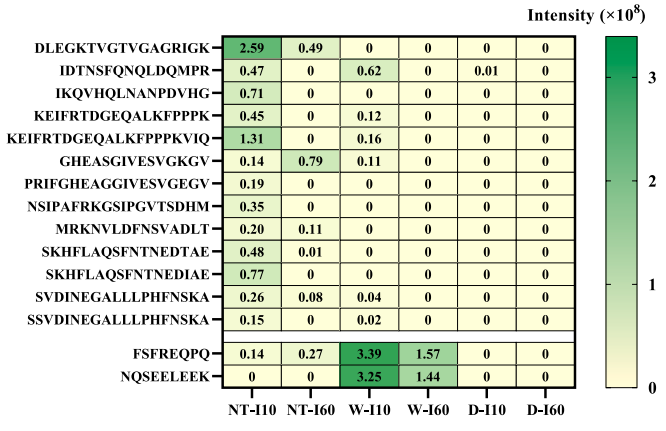

Fig. 4. Intensity of epitopic peptides in the digests under wet (W) or dry (D) heating, with non-treated (NT) samples as control. I10 and I60 represent 10 min and 60 min of intestinal digestion, respectively. Peptide sequences are displayed on the left side of their respective rows. The first 13 peptide sequences refer to the IL-4 inducing T-cell epitopes, while the last two sequences refer to the 100 % matching B-cell epitopes.

In terms of linear B-cell epitopes, after wet heating, both peptides "FSFREQPQ" and "NQSEELEEK" displayed an increasing trend in intensity. The intensities of both peptides increased by 3.3×10^8 at I10, starting from 1.4×10^7 and 0 (i.e., non-detectable), respectively. From I10 to I60, despite the initial increase in the intensity caused by wet heating, digestion subsequently decreased their intensities by 1.2- and 1.3-fold, respectively. In contrast, these peptides were non-detectable after dry heating at either I10 or I60. Therefore, unlike T-cell epitopes, B-cell epitopes exhibited an increase in intensity due to wet heating, and these epitopes were not completely eliminated during the following digestion. In other words, the immunogenic potential associated with B-cell epitopes eventually increased after wet heating, compared to the non-treated condition.

Regarding the mechanisms causing these differences, combining the findings from this study and from our earlier research (Tang et al., 2022, 2023), it is most likely that wet heating contributed to a looser tertiary structure resulting in increased accessibility to digestive enzymes, thus leading to two opposite results: the breakdown of the epitopic peptides, or the breakdown of longer peptide fragments into shorter epitopic peptides. For dry heating, the glycation-induced aggregation caused the formation of soluble aggregates and complicated cross-linked peptides after digestion, leading to the inability to detect single epitopic peptides. While for extended digestion, similar to wet heating, epitopic peptides were further digested and no longer existed in the digests, or conversely, longer peptides were hydrolyzed to certain peptides that possessed epitopic properties.

4. Conclusion

This study compared the soy protein-derived peptides released during *in vitro* infant digestion after wet and dry heating to the non-treated condition. The peptides exhibited a general decrease in total intensity and count from non-treated to wet-heated to dry-heated samples. Additionally, such decline was also found from 10 to 60 min of intestinal digestion. Subsequently, bioinformatics analysis predicted 29 potential T-cell epitopes among all peptides, with 13 having the potential to induce IL-4 secretion. Moreover, 27 possible linear B-cell epitopes were identified, with 2 perfectly matching previously reported B-cell epitopes. Furthermore, immunogenic cross-reactivity of putative epitopic peptides between soy proteins and other legumes including peanut, pea, chickpea, lentil, kidney bean, and lupine was explored. Most T-cell epitopic peptides decreased in intensity after either mode of heating,

while B-cell epitopes increased after wet heating but decreased after dry heating. After digestion, the intensities of these epitopic peptides eventually declined. However, the predicted epitopic peptides present at the initial stage of intestinal digestion (I10) may still trigger an immune response, even though they might be degraded later. These findings suggest that certain linear peptides released during digestion could potentially trigger immune responses, and heat treatment and digestion time could modulate this likelihood. Nevertheless, these results were only obtained by bioinformatics tools. Therefore, further validation through experiments is warranted, such as synthesizing these putative epitopes to conduct antibody binding assays and to evaluate their ability to activate T cells. Importantly, before making IF, understanding its immunogenic consequences is crucial. This study offers valuable insights into this area, thereby assisting in developing optimal plant-based formulas.

CRediT authorship contribution statement

Jiaying Tang: Conceptualization, Data curation, Formal analysis, Methodology, Software, Visualization, Writing – original draft. **Sjef Boeren:** Data curation, Formal analysis, Software, Writing – original draft. **Harry J. Wichers:** Conceptualization, Supervision, Writing – review & editing. **Kasper A. Hettinga:** Writing – review & editing, Validation, Supervision, Software, Funding acquisition, Conceptualization.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgment

This work was financially supported by the China Scholarship Council (No. 201906350087).

References

Aalberse, R. C. (2000). Structural biology of allergens. *Journal of Allergy and Clinical Immunology, 106*(2), 228–238. https://doi.org/10.1067/mai.2000.108434

Bai, T.-L., Han, X.-Y., Li, M.-S., Yang, Y., Liu, M., Ji, N.-R., & Liu, G.-M. (2021). Effects of the Maillard reaction on the epitopes and immunoreactivity of tropomyosin, a major allergen in Chlamys nobilis. *Food & Function, 12*(11), 5096–5108. https://doi.org/10.1039/D1FO00270H

Bublin, M., & Breiteneder, H. (2020). Cross-reactivities of non-homologous allergens. *Allergy, 75*(5), 1019–1022. https://doi.org/10.1111/all.14120

Cano, R. L. E., & Lopera, H. D. E. (2013). Introduction to T and B lymphocytes. In J. M. Anaya, Y. Shoenfeld, A. Rojas-Villarraga, R. A. Levy, & R. Cervera (Eds.), *Autoimmunity: From bench to bedside*. Bogota (Colombia): El Rosario University Press.

Cox, A. L., Eigenmann, P. A., & Sicherer, S. H. (2021). Clinical relevance of cross-reactivity in food allergy. *The Journal of Allergy and Clinical Immunology: In Practice, 9*(1), 82–99. https://doi.org/10.1016/j.jaip.2020.09.030

De Angelis, E., Pilolli, R., Bavaro, S. L., & Monaci, L. (2017). Insight into the gastro-duodenal digestion resistance of soybean proteins and potential implications for residual immunogenicity. *Food & Function, 8*(4), 1599–1610. https://doi.org/10.1039/C6FO01788F

Deng, Y., Govers, C., Bastiaan-Net, S., van der Hulst, N., Hettinga, K., & Wichers, H. J. (2019). Hydrophobicity and aggregation, but not glycation, are key determinants for uptake of thermally processed beta-lactoglobulin by THP-1 macrophages. *Food Research International, 120*, 102–113. https://doi.org/10.1016/j.foodres.2019.01.038

Dhanda, S. K., Gupta, S., Vir, P., & Raghava, G. P. S. (2013). Prediction of IL4 inducing peptides. *Clinical and Developmental Immunology, 2013*, Article 263952. https://doi.org/10.1155/2013/263952

Di Stasio, L., Picariello, G., Mongiello, M., Nocerino, R., Berni Canani, R., Bavaro, S., & Mamone, G. (2017). Peanut digestome: Identification of digestion resistant IgE binding peptides. *Food and Chemical Toxicology, 107*, 88–98. https://doi.org/10.1016/j.fct.2017.06.029

Dingess, K. A., de Waard, M., Boeren, S., Vervoort, J., Lambers, T. T., van Goudoever, J. B., & Hettinga, K. (2017). Human milk peptides differentiate between the preterm and term infant and across varying lactational stages. *Food & Function, 8*(10), 3769–3782. https://doi.org/10.1039/C7FO00539C

Han, X.-Y., Yang, H., Rao, S.-T., Liu, G.-Y., Hu, M.-J., Zeng, B.-C., & Liu, G.-M. (2018). The Maillard reaction reduced the sensitization of tropomyosin and arginine kinase from Scylla paramamosain, simultaneously. *Journal of Agricultural and Food Chemistry, 66*(11), 2934–2943. https://doi.org/10.1021/acs.jafc.7b05195

Hill, D. A., & Spergel, J. M. (2018). The atopic march: Critical evidence and clinical relevance. *Annals of Allergy, Asthma & Immunology, 120*(2), 131–137. https://doi.org/10.1016/j.anai.2017.10.037

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 (β-conglycinin) and Gly m 6 (glycinin) are potential diagnostic markers for severe allergic reactions to soy. *Journal of Allergy and Clinical Immunology, 123*(2), 452–458.e454. https://doi.org/10.1016/j.jaci.2008.09.034

Huang, Z., Qu, Y., Hua, X., Wang, F., Jia, X., & Yin, L. (2023). Recent advances in soybean protein processing technologies: A review of preparation, alterations in the conformational and functional properties. *International Journal of Biological Macromolecules, 248*, Article 125862. https://doi.org/10.1016/j.ijbiomac.2023.125862

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

Kamath, S. D., Bublin, M., Kitamura, K., Matsui, T., Ito, K., & Lopata, A. L. (2023). Cross-reactive epitopes and their role in food allergy. *Journal of Allergy and Clinical Immunology, 151*(5), 1178–1190. https://doi.org/10.1016/j.jaci.2022.12.827

Kattan, J. D., Cocco, R. R., & Järvinen, K. M. (2011). Milk and soy allergy. *Pediatric Clinics of North America, 58*(2), 407–426. https://doi.org/10.1016/j.pcl.2011.02.005

Klemola, T., Vanto, T., Juntunen-Backman, K., Kalimo, K., Korpela, R., & Varjonen, E. (2002). Allergy to soy formula and to extensively hydrolyzed whey formula in infants with cow's milk allergy: A prospective, randomized study with a follow-up to the age of 2 years. *The Journal of Pediatrics, 140*(2), 219–224. https://doi.org/10.1067/mpd.2002.121935

Knol, E. F., de Jong, N. W., Ulfman, L. H., & Tiemessen, M. M. (2019). Management of cow's milk allergy from an immunological perspective: What are the options? *Nutrients, 11*(11), 2734. https://doi.org/10.3390/nu11112734

L'Hocine, L., & Boye, J. I. (2007). Allergenicity of soybean: New developments in identification of allergenic proteins, cross-reactivities and hypoallergenization technologies. *Critical Reviews in Food Science and Nutrition, 47*(2), 127–143. https://doi.org/10.1080/10408390600626487

Leviatan, S., Vogl, T., Klompus, S., Kalka, I. N., Weinberger, A., & Segal, E. (2022). Allergenic food protein consumption is associated with systemic IgG antibody responses in non-allergic individuals. *Immunity, 55*(12), 2454–2469.e2456. https://doi.org/10.1016/j.immuni.2022.11.004

Li, H., Zhu, K., Zhou, H., Peng, W., & Guo, X. (2013). Comparative study about some physical properties, in vitro digestibility and immunoreactivity of soybean protein isolate for infant formula. *Plant Foods for Human Nutrition, 68*(2), 124–130. https://doi.org/10.1007/s11130-013-0358-9

Li, Y., Yang, Y., Zou, Y., Shu, L., Han, N.-Y., & Yang, Y. (2022). Characterization of the molecular properties and allergenicity (IgE-binding capacity) of β-lactoglobulin by heat treatment using asymmetric-flow field-flow fractionation and ultra-performance liquid chromatography quadrupole time of flight mass chromatography. *Food Chemistry, 374*, Article 131748. https://doi.org/10.1016/j.foodchem.2021.131748

Ménard, O., Bourlieu, C., De Oliveira, S. C., Dellarosa, N., Laghi, L., Carrière, F., & Deglaire, A. (2018). A first step towards a consensus static in vitro model for simulating full-term infant digestion. *Food Chemistry, 240*, 338–345. https://doi.org/10.1016/j.foodchem.2017.07.145

Moreno, M. d. L., Muñoz-Suano, A., López-Casado, M. Á., Torres, M. I., Sousa, C., & Cebolla, Á. (2016). Selective capture of most celiac immunogenic peptides from hydrolyzed gluten proteins. *Food Chemistry, 205*, 36-42. doi: 10.1016/j.foodchem.2016.02.066.

Mulet-Cabero, A.-I., Rigby, N. M., Brodkorb, A., & Mackie, A. R. (2017). Dairy food structures influence the rates of nutrient digestion through different in vitro gastric behaviour. *Food Hydrocolloids, 67*, 63–73. https://doi.org/10.1016/j.foodhyd.2016.12.039

Nutten, S., Maynard, F., Järvi, A., Rytz, A., Simons, P. J., Heine, R. G., & Kuslys, M. (2020). Peptide size profile and residual immunogenic milk protein or peptide content in extensively hydrolyzed infant formulas. *Allergy, 75*(6), 1446–1449. https://doi.org/10.1111/all.14098

Phongphisutthinant, R., Wiriyacharee, P., Boonyapranai, K., Ounjaijean, S., Taya, S., Pitchakarn, P., & Chaipoot, S. (2024). Effect of conventional humid-dry heating through the Maillard Reaction on chemical changes and enhancement of in vitro bioactivities from soy protein isolate hydrolysate-yeast cell extract conjugates. *Foods, 13*(3). https://doi.org/10.3390/foods13030380

Pi, X., Liu, J., Sun, Y., Ban, Q., Cheng, J., & Guo, M. (2023). Heat-induced changes in epitopes and IgE binding capacity of soybean protein isolate. *Food Chemistry, 405*, Article 134830. https://doi.org/10.1016/j.foodchem.2022.134830

Sanchez-Trincado, J. L., Gomez-Perosanz, M., & Reche, P. A. (2017). Fundamentals and methods for T- and B-cell epitope prediction. *Journal of Immunology Research, 2017*(1), 1–14. https://doi.org/10.1155/2017/2680160

Singh, H., Ansari, H. R., & Raghava, G. P. S. (2013). Improved method for linear B-cell epitope prediction using antigen's primary sequence. *Plos One, 8*(5), e62216.

Tang, J., Wichers, H. J., & Hettinga, K. A. (2022). Heat-induced unfolding facilitates plant protein digestibility during in vitro static infant digestion. *Food Chemistry, 375*, Article 131878. https://doi.org/10.1016/j.foodchem.2021.131878

Tang, J., Wichers, H. J., & Hettinga, K. A. (2023). Glycation of soy and pea proteins influences infant gastric digestibility more than intestinal digestibility. *Food Hydrocolloids, 136*, Article 108251. https://doi.org/10.1016/j.foodhyd.2022.108251

Van Wijk, F., Hartgring, S., Koppelman, S. J., Pieters, R., & Knippels, L. M. J. (2004). Mixed antibody and T cell responses to peanut and the peanut allergens Ara h 1, Ara h 2, Ara h 3 and Ara h 6 in an oral sensitization model. *Clinical & Experimental Allergy, 34*(9), 1422–1428. https://doi.org/10.1111/j.1365-2222.2004.02062.x

Verma, A. K., Kumar, S., Das, M., & Dwivedi, P. D. (2013). A comprehensive review of legume allergy. *Clinical Reviews in Allergy & Immunology, 45*(1), 30–46. https://doi.org/10.1007/s12016-012-8310-6

Vita, R., Mahajan, S., Overton, J. A., Dhanda, S. K., Martini, S., Cantrell, J. R., & Peters, B. (2018). The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Research, 47*(D1), D339–D343. https://doi.org/10.1093/nar/gky1006

Wang, C., Wang, Y., Liu, G., & Fu, L. (2020). Food allergomics based on high-throughput and bioinformatics technologies. *Food Research International, 130*, Article 108942. https://doi.org/10.1016/j.foodres.2019.108942

Xiong, L., Boeren, S., Vervoort, J., & Hettinga, K. (2021). Effect of milk serum proteins on aggregation, bacteriostatic activity and digestion of lactoferrin after heat treatment. *Food Chemistry, 337*. https://doi.org/10.1016/j.foodchem.2020.127973

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. https://doi.org/10.1016/S0022-3476(99)70249-0

Zenker, H. E., Wichers, H. J., Tomassen, M. M. M., Boeren, S., De Jong, N. W., & Hettinga, K. A. (2020). Peptide release after simulated infant in vitro digestion of dry heated cow's milk protein and transport of potentially immunoreactive peptides across the Caco-2 cell monolayer. *Nutrients, 12*(8), 2483. https://doi.org/10.3390/nu12082483

Zhou, F., He, S., Zhang, Y., Wang, Y., Sun, H., & Liu, Q. (2022). Prediction and characterization of the T cell epitopes for the major soybean protein allergens using bioinformatics approaches. *Proteins: Structure, Function, and Bioinformatics, 90*(2), 418–434. https://doi.org/10.1002/prot.26233