Prediction and characterization of the T cell epitopes for the major soybean protein allergens using bioinformatics approaches

Fanlin Zhou[1], Shudong He[1], Yongfei Wang[1], Hanju Sun[1], and Yi Zhang[2]

[1]Hefei University of Technology
[2]McGill University

July 22, 2021

Abstract

Protein allergens is a health risk for consumption of soybeans. To understand allerginicity mechanism, T cell epitopes of 7 soybean allergens were predicted and screened by abilities to induce cytokine interleukin 4. The relationships among amino acid composition, properties, allergenicity and pepsin hydrolysis sites were analyzed. Among the 138 T cell epitopes identified, YIKDVFRVIPSEVLS, KDVFRVIPSEVLSNS, DVFRVIPSEVLSNSY of Gly m 6.0501 (P04347), and AKADALFKAIEAYLL, ADALFKAIEAYLLAH of Gly m 4.0101 (P26987) were the most possible epitope candidates. In T cell epitopes pattern, the frequencies of amino acids Q, D, E, P and G decreased, while F, I, N, V, K and H increased. Hydrophobic residues at positions p1 and p2 and positively charged residues in positions p13 might contribute to allergenicity. Most of epitopes could be hydrolyzed by pepsin into small polypeptides within 12 residues length, and the anti-digestive epitope regions contained I, V, S, N, and Q residues. T cell epitopes EEQRQQEGVIVELSK from Gly m 5.03 (P25974) showed resistantence to pepsin hydrolysis and would cause a higher Th2 cell response. This research provides basis for the development of hypoallergenic soybean products in the soybean industry as well as for the immunotherapy design for protein allergy.

1. Introduction

In recent years, the incidence of food allergies has been on the rise, and about 10% of people in the world are suffering from food allergies [1]. So far, 390 proteins have been identified as food protein allergens by the World Health Organization and the International Union of Immunological Societies (WHO/IUIS) Allergen Nomenclature Sub-Committee, and more than 50% of the food protein allergens are found in eggs, milk, nuts, wheat, crustaceans, beans, fish, and peanuts [2]. Soybean is a common source for legume allergen, and about 25% of food allergies are caused by soybeans [3, 4], which not only induce a variety of pathological reactions such as intestinal injury, stomach discomfort, or allergic dermatitis, anaphylactic shock, [3], but also limit the development and application of soybean products [4]. So far, eight soybean protein allergens have been identified and submitted to the allergen database (http://www.allergen.org/), namely hydrophobic protein (Gly m 1), defensin (Gly m 2), profilin (Gly m 3), pathogenesis-related protein (Gly m 4), β-conglycinin (Gly m 5), glycinin (Gly m 6), seed biotinylated protein (Gly m 7) and 2S albumin (Gly m 8). Besides, Gly m Bd 28k, Gly m Bd 30k, trypsin inhibitor, lectin, and Gly 50 kDa have been also confirmed as the common soybean allergens [5]. However, among these soybean protein allergens, only seven of them, i.e., Gly m 4.0101 (Uniprot ID: P26987), Gly m 5.0201 (Uniprot ID: P11827), Gly m 5.03 (Uniprot ID: P25974), Gly m 6.0101 (Uniprot ID: P04776), Gly m 6.0501 (Uniprot ID: P04347), trypsin inhibitor (Uniprot ID: P01070) and lectin (Uniprot ID: P05046) (Table 1), have been characterized with crystal structures as seen in the Uniprot database (https://www.uniprot.org/).

Previous studies have indicated that soybean allergy is usually mediated by type 2 CD4+ T cells (Th2) [5]. In detail, the T cell epitopes of soybean allergens combine with major histocompatibility complex (MHC) class II proteins, which can be recognized by Th2 cells and induce the release of IL-4, IL-5, IL-13 and

other interleukins that promote B cell proliferation and differentiation to produce antibodies, and play an important role in activating mast cells, basophils, and eosinophils [2]. Since antigen recognition by T cells is a critical step, the analyses of T cell epitopes and related cytokines are of vital importance for understanding the mechanism of anaphylaxis. Generally, T cell epitopes typically consist of 12-20 amino acids, and the T cell epitope chain less than 12 amino acids would not efficiently stimulate CD4 cells [6]. For the peptide fragment that are bound by MHC class II molecules, they are usually longer (13-15 amino acids) where the core sequences have length of 9 amino acids [7]. MHC molecules are known as peptide-binding glycoproteins, a complex that affect the immune response. For major histocompatibility antigen system in human body, the histocompatibility leukocyte antigens (HLAs) encoded by the MHC genes produce the Class I and class II molecules, and HLA class II molecules (HLA-DP, HLA-DQ, HLA-DR) alleles are gene complexes encoding human MHC class II proteins [8]. However, the alleles of HLA class II alleles are not clear, and the epitope peptides fragment bound to HLA class II molecules as well as the binding affinity are rarely described.

The identification of T cell epitopes of food allergens helps with the understanding of the allergy mechanisms and the development of hypoallergenic foods and immunotherapy design. As previously reported, employing appropriate proteases to pre-hydrolyze food allergens are effective to disrupt the allergenicity of food proteins, but the protease enzymes need to be selecgted based on the anti-digestion area of epitopes [9]. For example, there are many hypoallergenic children's formula milk products that are developed via enzymatic pre-hydrolysis method, in order to effectively prevent children from milk allergy, and improve taste and nutritional value [10]. Furthermore, T cell epitopes have been considered as safe and effective immunotherapy modulators [2]. For example, the allergic reactions from shrimp tropomyosin and egg ovalbumin studied on mice could be effectively reduced by T cell epitopes oral immunotherapy [11, 12]. The knowledge of allergen-specific T cell epitopes is also useful to develop T cell epitope targeted vaccines for immunotherapy [13].

Allergen T cell epitopes are conventionally identified using peptide scanning technology via synthesizing multiple overlapping peptides that span the length of the protein in the combination with *in vitro* T cell stimulation tests, but this approach was costly in the peptide synthesis and serum preparation [14]. Although the lymphocyte proliferation assay in the T cell test is simple, inexpensive and sensitive, it takes long time and can not identify and quantify the cytokines. Flow cytometry is a novel T cell epitope identification method with high throughput screening to save time, but the equipment is expensive [15]. In this context, bioinformatics has been applied as an initial screening methoda and widely used in the prediction of food allergen epitopes with advantages including high throughput, fast analysis speed, and low cost, benefiting from database development [16]. In our previous studies, seven different bioinformatics tools, ProPred, SYFPEITHI, NetMHCII, NN-align, SMM-align, NetMHCIIpan and RANKPEP, were employed to predict the T cell epitopes of a black kidney bean, from which two of three potential T cell epitopes were confirmed using cytokine and lymphocyte proliferation experiments [17], suggesting the prediction accuracy of bioinformatics analysis. Besides, thirty-six T cell epitopes of the peanut allergen Ara h 1 were obtained by NetMHCIIpan 2.0 tool, and 14 of predicted T cell epitopes binded to HLA molecules *in vitro* , and 35 of T cell epitopes induced T cell proliferation differentiation and the production of IL13 [14]. By now, only one soybean allergen, Gly m 6, with identified epitopes, is recorded in SDAP database, and most of the relevant studies pay attention to the exploration of B cell epitopes from soybean allergens. For instance, Zeece et al. [18] identified the IgE epitope (aa192-306) in the region of soybean globulin G1 acidic subunit by western blot. Helm et al. [19] used the peptide scanning technique to identify 11 linear B cell epitopes of soybean globulin G1, and the epitopes AGVALSRCTLN (aa 62-72) had cross-reactivity with a peanut allergen. Saeed et al. [20] identified 9 glycinin epitopes and found that glycinin cross-reacted with peanuts, almonds, and walnuts. However, there is a lack of research on the T cell epitopes of soybean allergens.

In this study, in order to make a comprehensive understanding of T cell epitopes of soybean allergens, all the seven soybean protein allergens characterized with crystal structures were subjected for bioinformatic analyses. T cell epitopes of soybean allergens were predicted by*in silico* tools, and the binding affinity with different HLA II alleles and the ability to induce IL-4 as well as the allergenicity of epitopes were also evaluated to understand the allergenicity mechanisms. Amino acid composition and the quantitative

structure - activity relationship as well as the pepsin hydrolysis sites of epitopes were analyzed.

2.

Methods

2.1 Allergen amino acid sequence retrieval

The amino acid sequences of P01070 (trypsin inhibitor), P04347 (Gly m 6.0501), P04776 (Gly m 6.0101), P05046 (lectin), P11827 (Gly m 5.0201), P25974 (Gly m 5.03), and P26987 (Gly m 4) were downloaded from the Uniprot database (https://www.uniprot.org/), and the detailed information of the seven soy allergens are shown in Table 1.

2.2 T cell epitope prediction

The "MHC-II Binding Predictions" tool in the IEDB (Immune Epitope Database Analysis Resource) database was used to predict the soybean allergen peptides that could bind to HLA class II molecules [21]. The potential sequence was submitted to the software in Fasta format, and the IEDB recommended method, combining with the consensus method and the NetMHCIIpan method, was selected for prediction, and thereinto, the consensus method considers the combination of any three of four methods, including artificial neural network (ANN) alignment method, stabilization matrix (SMM) alignment method, combinatorial library method, and Sturniol method. A total of 27 HLA molecules (15 HLA-DR molecules, 6 HLA-DQ molecules, and 6 HLA-DP molecules) were used to predict T cell epitopes, and the epitope length parameter was set to 15 amino acids as a suggestion.

2.3 Evaluation of the ability to induce IL-4

The IL4pred tool (https://webs.iiitd.edu.in/raghava/il4pred/) was used to calculate the ability of epitopes to induce IL-4 secretion by Th2 cells, and a hybrid prediction method consisting of support vector machine and motif was selected. The support vector machine threshold was selected as 0.2 [22].

2.4 Allergenicity evaluation

The amino acid sequences of the T cell epitopes were submitted to the online tool AllerTop 2.0 (https://www.ddg-pharmfac.net/AllerTOP/), and the amino acid properties, such as hydrophobicity, size, helix-forming propensity, relative abundance, and β-strand forming propensity, were applied for the allergenicity prediction [23, 24].

2.5 Epitope amino acid composition analysis

The amino acid composition of T cell epitopes and the frequency were analyzed according to the methods of He et al. [17] and Zheng et al. [25].

2.6 Significance analysis of epitope amino acids

The amino acid descriptor 3z scales (z1 descriptor represents hydrophilicity, z2 represents the bulk, and z3 represents electricity properties) were used as the X variable and the T cell epitope allergenicity (yes/no) was set as the Y variable, [26]. Each T cell epitope had 15 aa and the number of variables for each soybean allergen was 45 (3*15), expressed a,s p1z1-p15z3. The discriminant analysis was performed with the XLSTAT software (Trial version, Addinsoft, New York, NY) to confirm the accuracy of the amino acid descriptor 3z scales. then the significance analysis of T cell epitope amino acids was carried out by using the logistic regression and random forests via the Easy Fit function of XLSTAT software.

2.7 Simulated pepsin digestion of epitopes

The online tool PeptideCutter (https://web.expasy.org/peptide_cutter/) was used to predict the pepsin digestion sites of the T cell epitopes of soybean allergens [27].

3. Results

3.1T cell epitope prediction

In order to compare the binding affinity between different peptides and HLA class II molecules, a unified scale has been generated in the prediction results of the IEDB-percentile ranking value, as the lower percentile ranking value with higher affinity [28]. In addition, the inhibitory concentration 50 (IC50) value was also used to calculate the peptide binding affinity, as the lower, IC50 value indicated the stronger affinity [24]. Therefore, peptides with a strong binding ability to MHC class II molecules would be screened out based on the peptide binding affinity score with IC50[?]250 nM and percentile ranking value[?]4. Subsequently, T cell peptides with percentile ranking value [?] 4, stable matrix method IC50 [?] 250 nM, and neural network method IC50 [?] 250 nM obtained by the consensus method, and the peptides with the percentile ranking value [?] 4, and NetMHCIIpan IC50 [?] 250 nM obtained by the NetMHCIIpan method were further selected for the IL-4 inducing ability analysis, and the AllerTop 2.0 tool was finally used to confirm the allergenicity potential.

3.1.1 P01070

Soybean trypsin inhibitor (Uniprot ID: P01070) is an anti-nutritional factor in soybean [29], consisting of 216 amino acids, with a protein molecular weight of 24005 Da. Forty-five peptides with the high binding ability to MHC class II molecules were obtained by setting IC50 [?] 250 nM and percentile ranking value [?] 4, and 22 peptides could induce IL-4 secretion by Th2 cells based on the IL4pred tool analysis. Furthermore, a total of 14 peptides were confirmed as the T cell epitopes via the allergenicity analysis using the AllerTOP v. 2.0 tool (Table 2), which were mainly located in three regions of the protein: aa 44-60, aa 86-111, and aa 179-199. The region 44-60 are located inside the protein, whereas segments aa 86-111, and aa 179-199 were exposed to the protein surface (Table 2). Thereinto, the epitopes "YRIRFIAEGHPLSLK" (aa 86-100) and "RIRFIAEGHPLSLKF" (aa 87-101) might have a higher ability to induce Th2 to produce IL4 because of the higher IL-4pred scores. Furthermore, as shown in Table 2, the bindings of the epitopes "PLSLKFDSFAVIMLC" (aa 96-110) and "LSLKFDSFAVIMLCV" (aa 97-111) with the most diverse HLA class II alleles were observed, indicating these two epitopes can cause more people to be allergic than other epitopes.

3.1.2 P04347

Gly m 6.0101, Gly m 6.0201, Gly m 6.031, Gly m 6.0401, and Gly m 6.0501 are the 5 subunits (G1-G5) of the Glycinin, which belong to the 11S plant seed storage protein [30]. Gly m 6.0501 (Uniprot code: P04347) containing 516 amino acids with a molecular weight of 57956 Da was analyzed in the study, and a total of 22 T cell epitopes were screened (Table 3), which were mainly located in 5 protein regions of aa 183-199, aa 231-246, aa 260-275, aa 390-407, and aa 461-495. Thereinto, protein fragments of aa 231-246, aa 260-275, and aa 471-494 are posited on the surface of the protein (Table 3). Compared with other epitopes, the four epitopes of "GLEYVVFKTHHNAVS" (aa 461-475), "LEYVVFKTHHNAVSS" (aa 462-476), "YVVFKTHHNAVSSYI" (aa 464-478), and "VVFKTHHNAVSSYIK" (aa 465-479) had higher IL-4pred scores.

3.1.3 P04776

Gly m 6.0101 (Uniprot ID: P04776) is the G1 subunit of glycinin, which contains 495 amino acids and has a molecular weight of 55706 Da. As shown in Table 4, a total of 34 T cell epitopes, mainly concentrated in the 7 regions of aa 158-172, aa 217-236, aa 319-333, aa 347-362, aa 366-387, aa 412-442 and aa 468-491, that could induce Th2 cell to produce IL4 and have allergenicity potential. Among them, the regions of aa 217-236, aa 319-333, and aa 468-491 are exposed on the surface (Table 4). T cell epitopes in the aa 217-236 region could bind the largest number of HLA class II alleles, indicating the allergic susceptibility, and the epitopes "ILSGFTLEFLEHAFS" (aa 220-234) and "LSGFTLEFLEHAFSV" (aa 221-235) had higher IL-4pred scores.

3.1.4 P05046

Soybean lectin (Uniprot ID: P05046) is an anti-nutritional factor with 285 amino acids and the molecular weight is approximately 120 kDa [31]. As shown in Table 5, a total of 25 peptides were confirmed as

the T cell epitopes that can induce Th2 to produce IL4 and have allergenicity potential. All T cell epitopes are located on the surface of the protein except for the region aa 234-250 (Table 5). Furthermore, the epitopes "LVLLTSKANSAETVS" (aa 23-37), "EWVRIGFSAATGLDI" (aa 234-248), "VRIGFSAATGLDIPG" (aa 236-250), "HDVLSWSFASNLPHA" (aa 253-267), "DVLSWSFASNLPHAS" (aa 254-268) and "VLSWSFASNLPHASS" (aa 255-269) exhibited higher IL-4pred scores. Especially, the epitope "ASFAASFNFTFYAPD" (aa 100-114) was predicted by three methods at the same time, and more HLA class II alleles were observed to be bound by this epitope.

3.1.5 P11827

Soybean β-conglycinin is a 7S seed storage protein containing three subunits (α, α'and β) [32]. Gly m 5.0201 (Uniprot ID: P11827) is the α' subunit of β-conglycinin and has high amino acid sequence homology with the α and β subunits, and the molecular weight is 72,228 Da containing 621 amino acids [33]. Through IEDB tools, IL-4pred, and AllerTOP v. 2.0 prediction, a total of 17 T cell epitopes were screened (Table 6), and the epitopes "PFHFNSKRFQTLFKN" (aa 211-225) and "FHFNSKRFQTLFKNQ" (aa 212-226) combined the most types of alleles, and they also had strong potentials for inducing Th2 cells to produce IL4. Also, these two epitopes are exposed on the surface of the protein (Table 6).

3.1.6 P25974

Soybean β-conglycinin β subunit (Uniprot code: P25974) is composed of 439 amino acids and has a molecular weight of 50476 Da. A total of 15 T cell epitopes were obtained with allergic potential (Table 7). Compared with other epitopes, the epitopes "EEQRQQEGVIVELSK" (aa 201-215), "EQRQQEGVIVELSKE" (aa 202-216), and "RNPIYSNNFGKFFEI" (aa 245-259) had higher IL-4pred scores and were located on the surface of the protein.

3.1.7 P26987

Gly m 4 (Uniprot ID: P26987) is the pathogenesis-related protein that belongs to the PR protein family. It has cross-reactivity with apple allergen Mal d 1, birch pollen allergen Bet v 1, and other allergens [34]. Gly m 4 has a protein length of 158 amino acids and a molecular weight of 16772 Da, while a total of 12 T cell epitopes were screened by the prediction method (Table 8). These 12 T cell epitopes are all posited on the surface of the protein. Compared with other epitopes, the epitopes "AKADALFKAIEAYLL" (aa 138-152) and "ADALFKAIEAYLLAH" (aa 140-154) bond the most types of HLA class II alleles, 13 and 15 different HLA alleles, respectively.

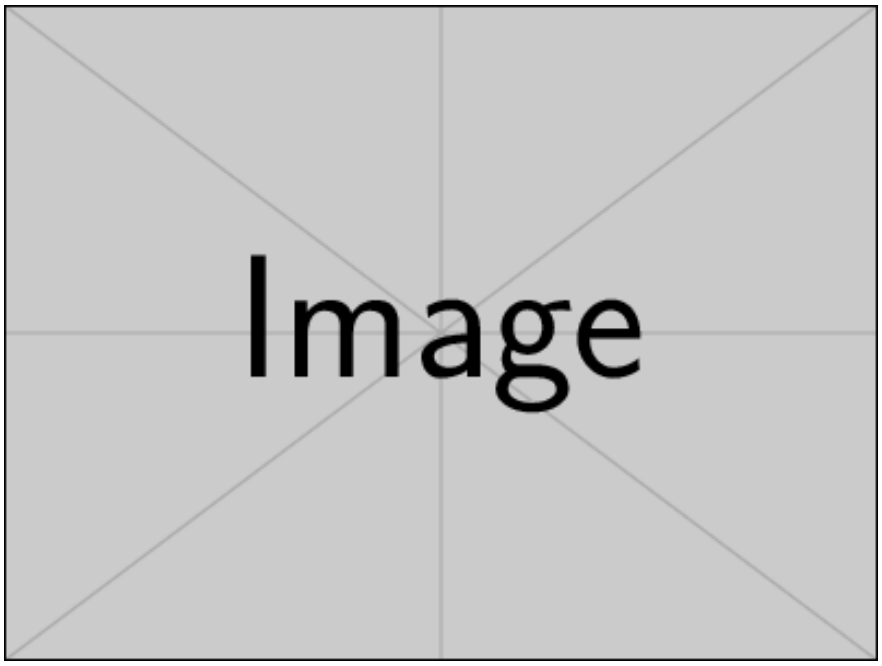
3.2 Amino acid composition of T cell epitopes

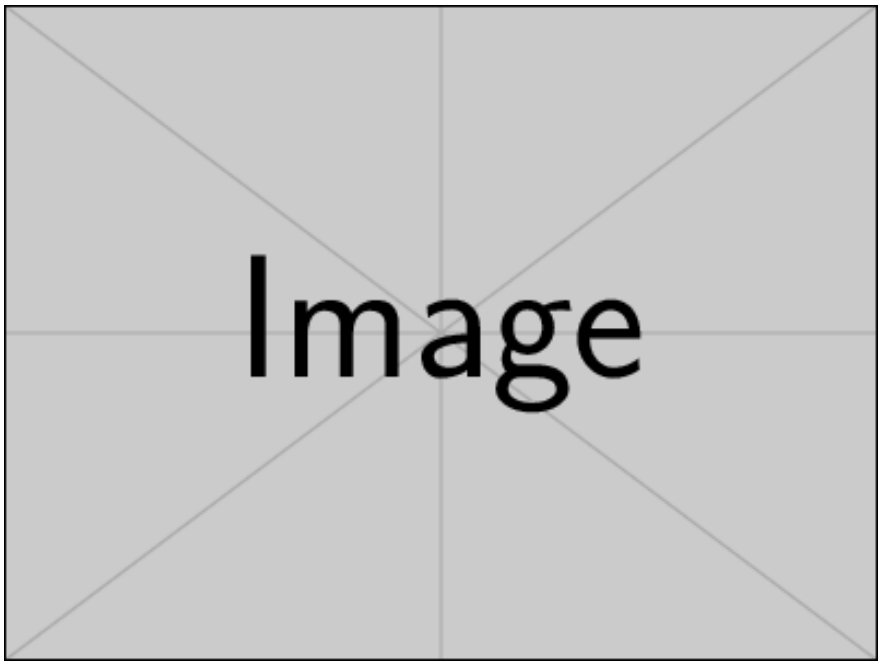
As shown in Fig. 1, the ratio of each amino acid in the entire protein and that in the T cell epitope fragment were counted, respectively. All 20 types of amino acids were observed in four of the seven proteins (P01070, P04347, P04776, and P11827), and the P05046 protein and P25974 protein contained 19 types of different amino acids except for cysteine (C) and tryptophan (W), respectively, while 17 types of amino acids were found in the P26987 protein except for arginine (R), tryptophan (W), and cysteine (C). While lysine (L), serine (S), valine (V), glutamate (E), and asparagine (N) were found to be rich in the seven protein allergens, and fewer tryptophane (W), methionine (M) and cysteine (C) were observed.

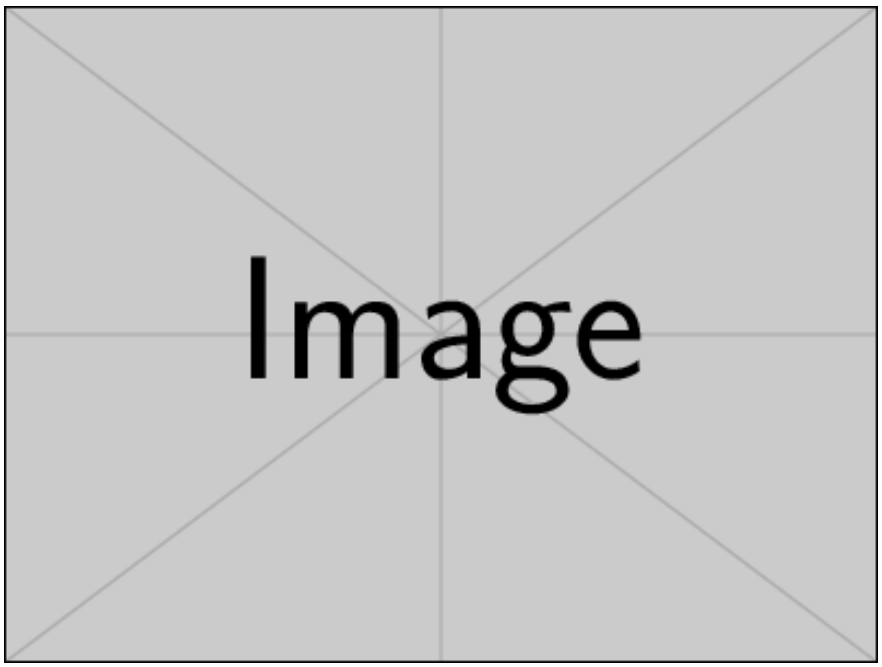
By comparing the frequency of each amino acid in the protein sequence with that in the T cell epitopes, the frequency of glutamine (Q) is obviously reduced in the T cell epitope region of the six proteins (except for P25975 protein). The appearance frequency of glutamate (E) in the T cell epitope regions of P01070, P04347, P04776, P05046, and P11827 proteins was also decreased, and a similar phenomenon was also observed in the T cell epitope region of P04347, P05046, P11827, P25974, and P26987 proteins for proline (P), while the frequency of glycine (G) in T cell epitope regions was reduced in P04347, P04776, P11827 and P26987 proteins, and aspartate (D) was decreased in P01070, P04347, P05046, and P25974 proteins. On the other hand, the frequency of phenylalanine (F) in T cell epitope regions was increased among 6 proteins (except for P01070 protein). In addition, the frequency of isoleucine (I), asparagine (N), valine (V), lysine (K), and histidine (H) in T cell epitope regions were also increased in more than half of the proteins (4/7).

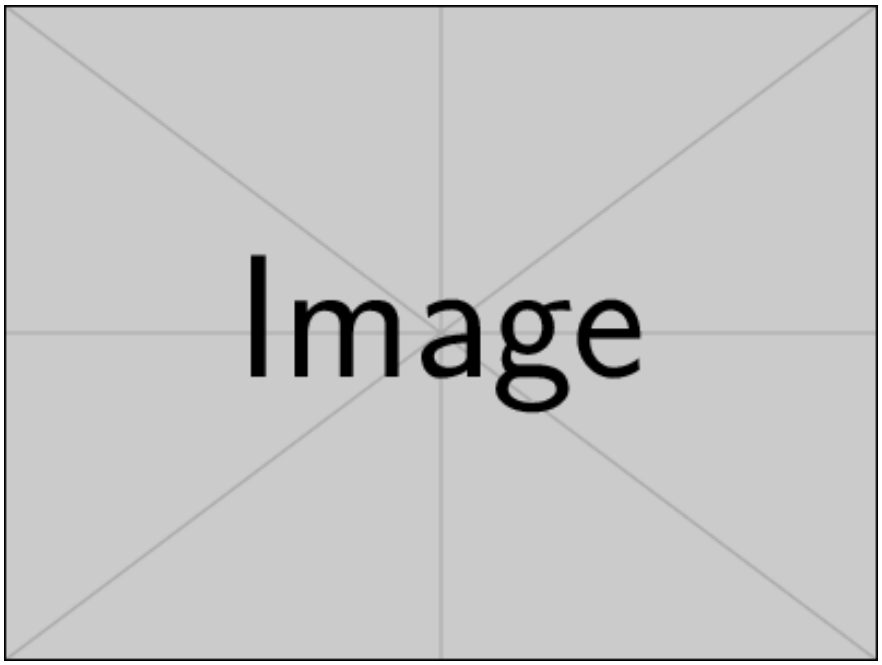
Hosted file

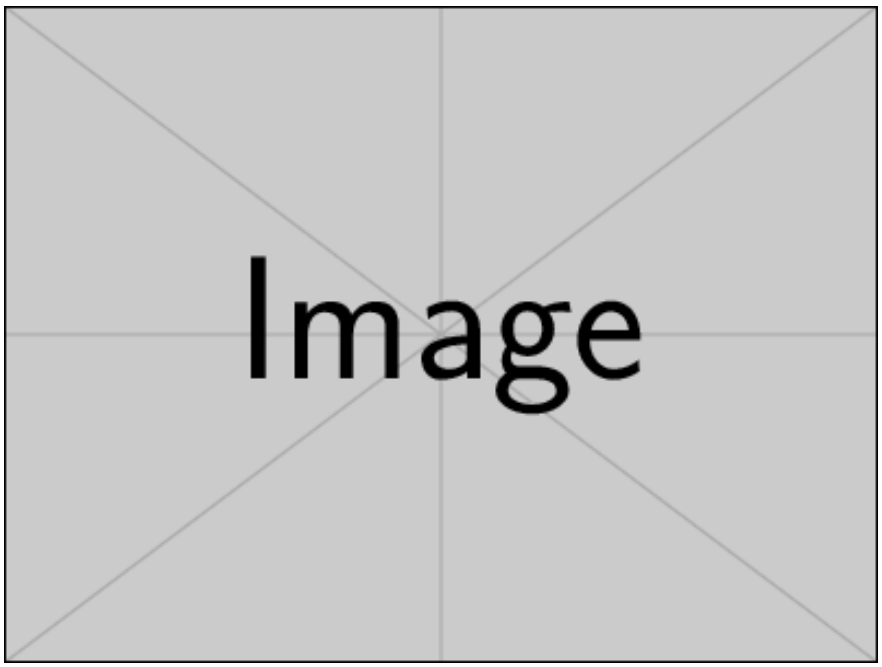
`image1.emf` available at https://authorea.com/users/427031/articles/531346-prediction-and-characterization-of-the-t-cell-epitopes-for-the-major-soybean-protein-allergens-using-bioinformatics-approaches

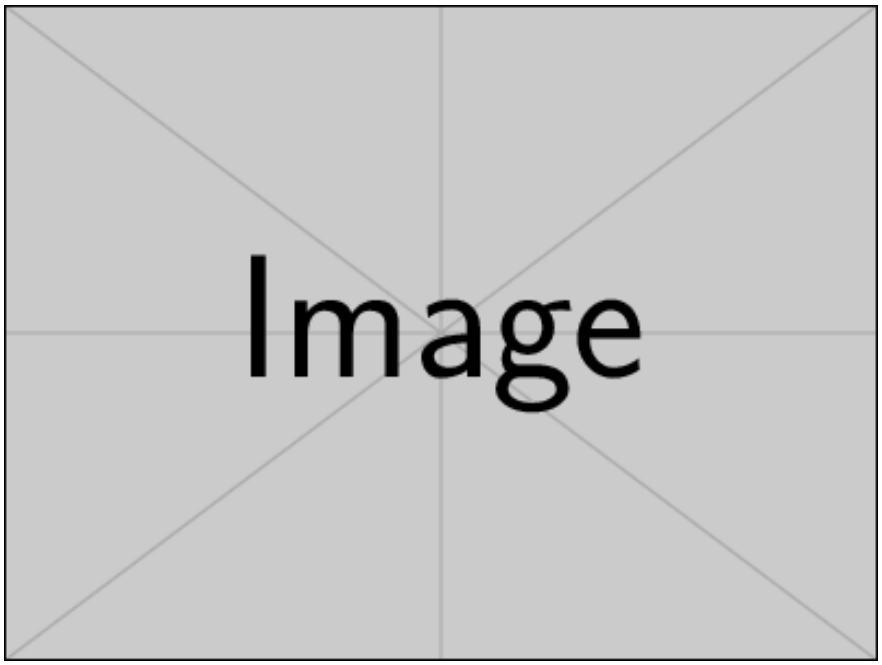


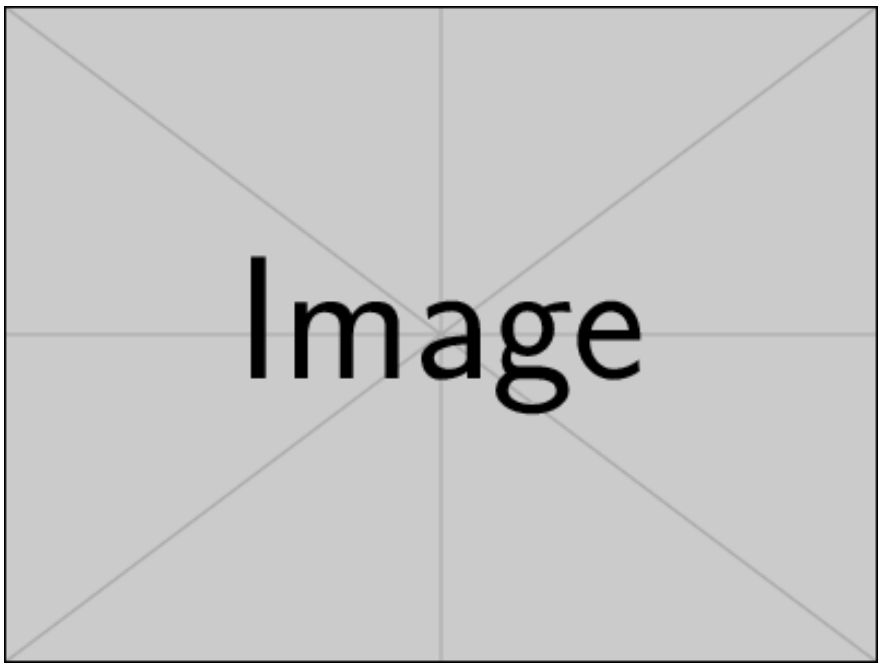


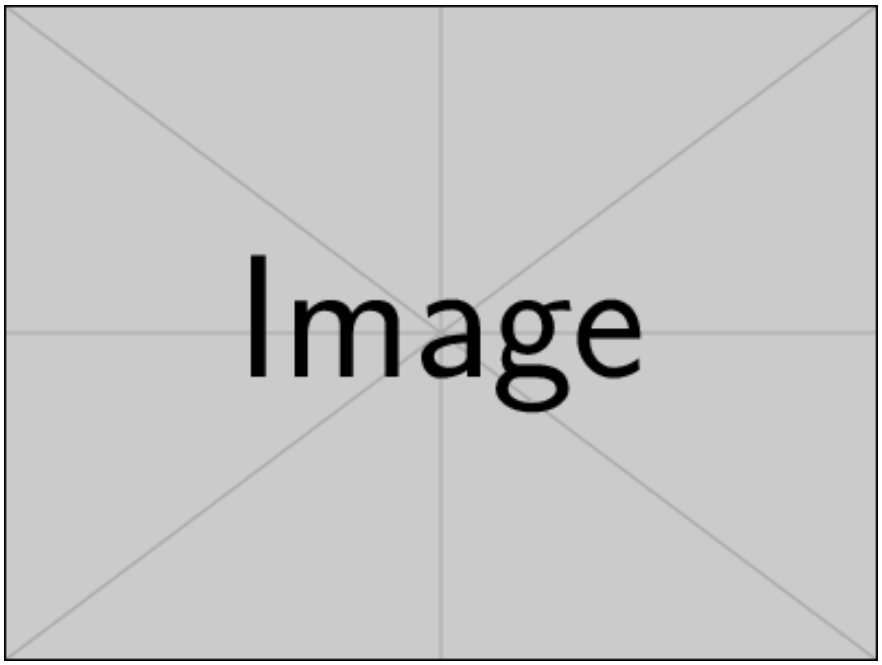


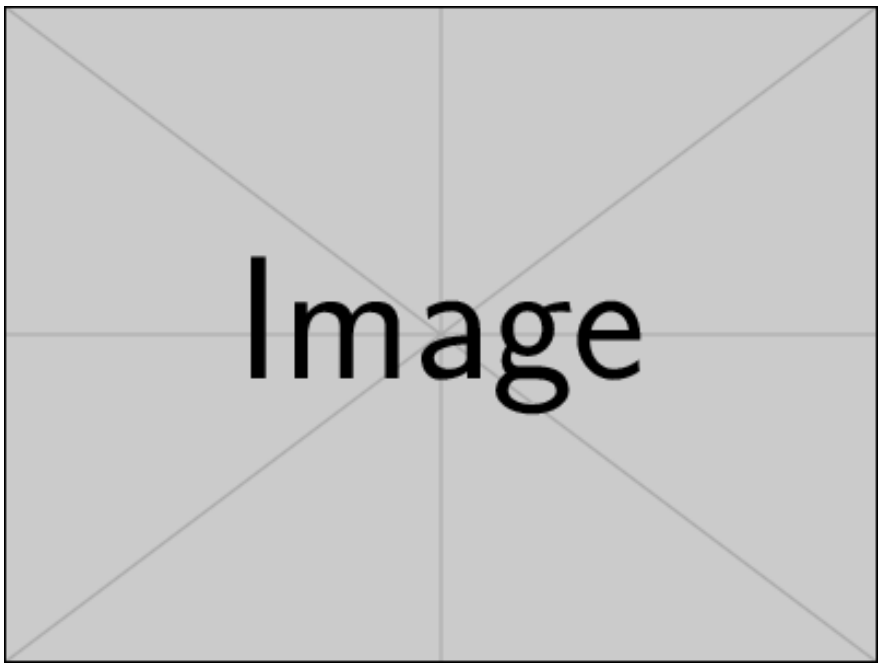


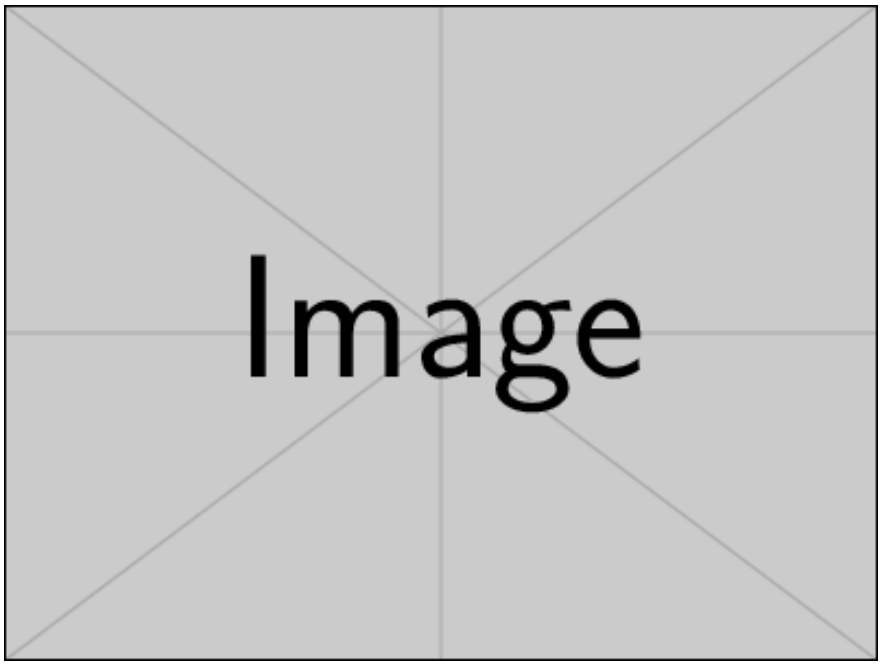


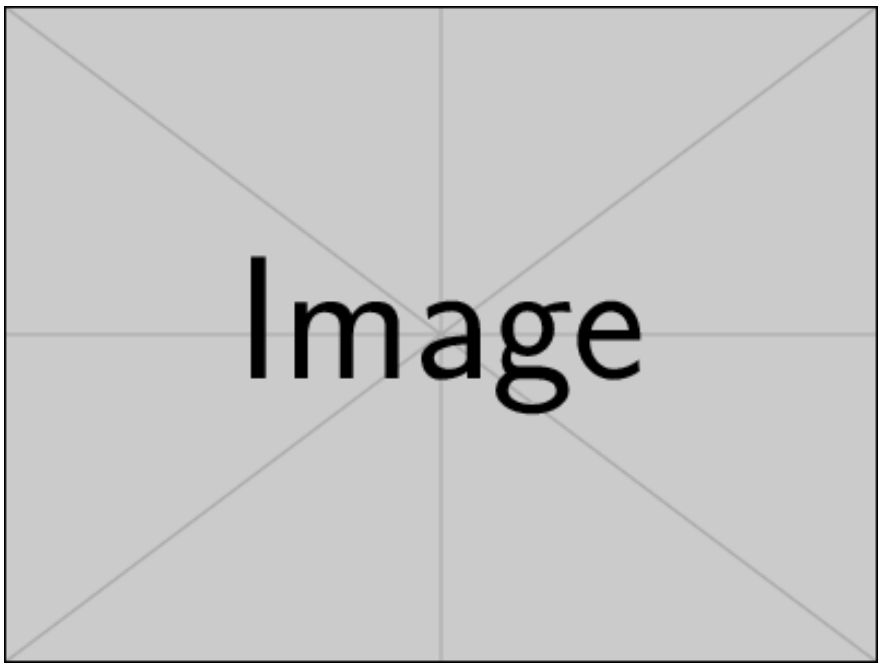


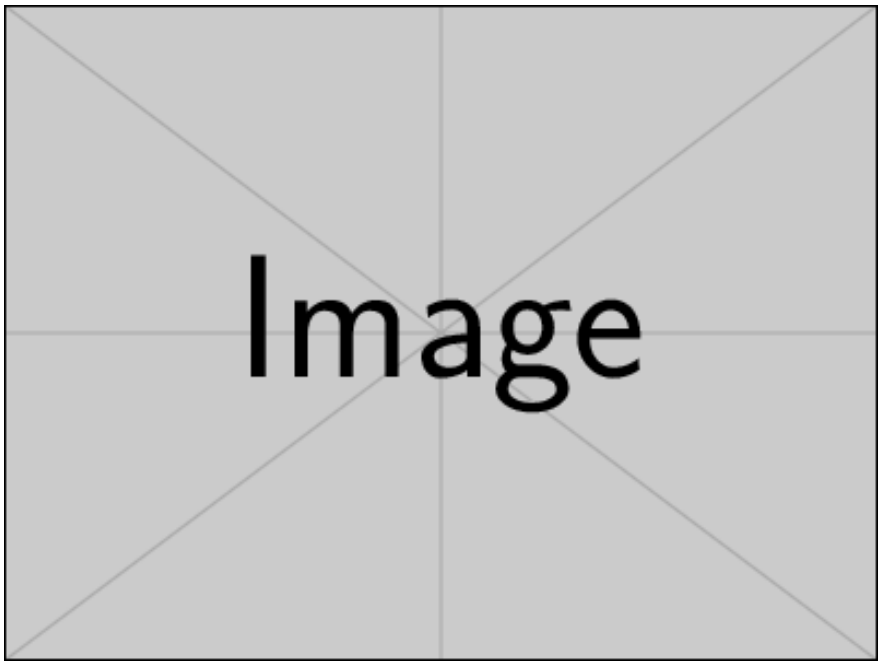


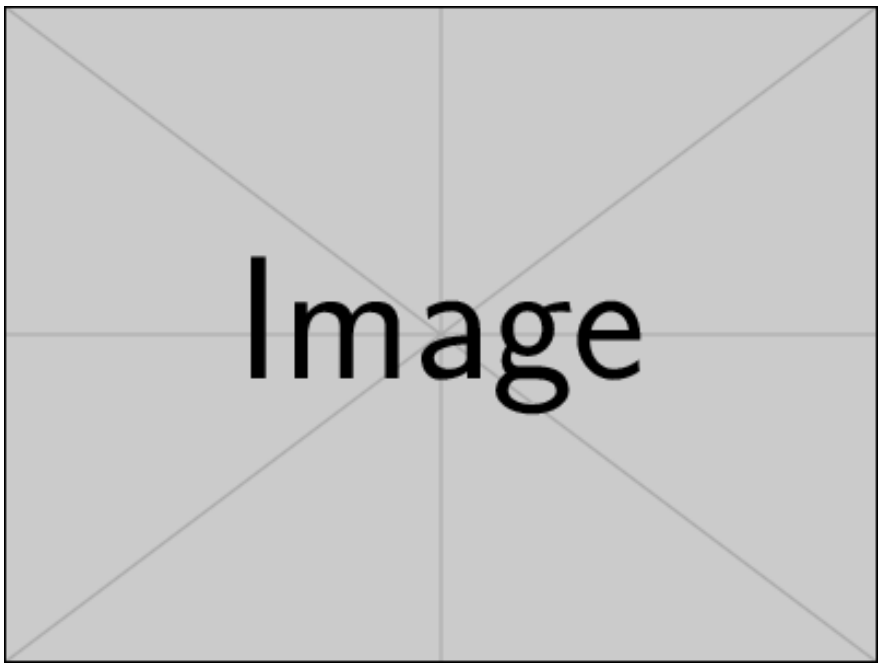


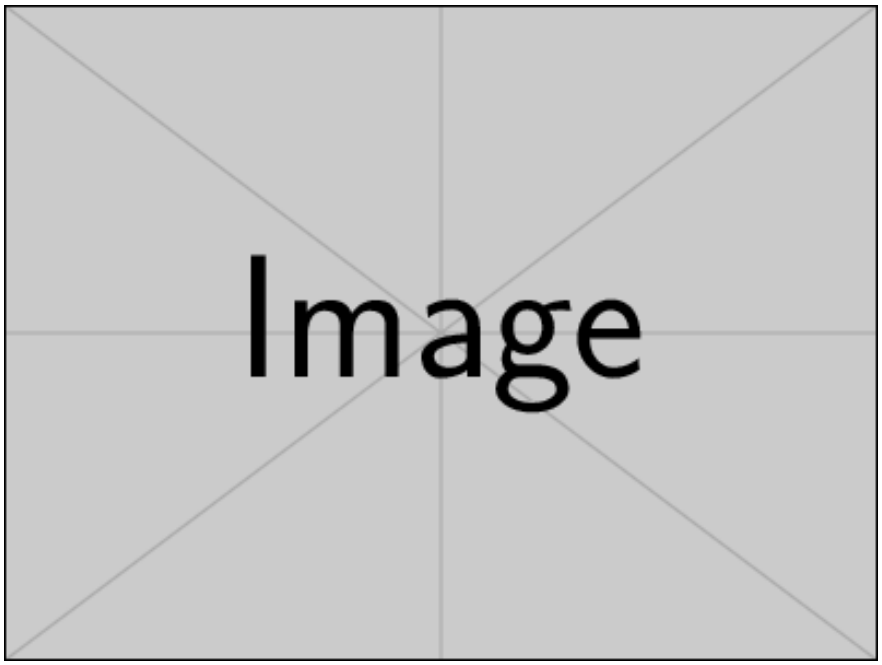












3.3 Significance analysis of epitope amino acids

A discriminant analysis model was used to analyze the relationship between the allergenicity (yes or no, Y-variable) of the epitope peptides and the physical and chemical properties of the amino acids (X- variables).

The present study used 3 descriptors to describe the amino acid physical and chemical properties of the T cell epitopes for building the random forests models with variable importance analysis. The confusion matrix (validation sample) correct (%) of seven soybean allergens were: 40% (P01070), 57.143% (P04347),

41.667% (P04776), 55.556% (P05406), 72.727% (P11827), 60% (P25974), and 75% (P26987). The variable importance of the X-variables was determined by examining the mean decrease accuracy obtained through random forests analysis of the quantitative X- variables and qualitative Y-variable (Fig. 2).

As shown in Fig. 2, the variable p1z1 significantly contributed to the allergenicity (yes) in the five soybean allergens (P01070, P04776, P05406, P11827, P25974). According to mean decrease accuracy values, variables p2z1, p6z2, p13z3 were beneficial to the allergenicity (yes) in four soybean allergens. Through the calculation, the occurrence of allergenicity (yes), p1, and p6 was the most important position to allergenicity (yes), followed by p2, p4, p5, and p13. The soybean allergens P01070, P11827, and P25974 expressed the bulk of the amino acid at the p1 position, whereas P04776 and P05046 expressed the electronic property of the amino acid at the p1 position. The amino acids at the p6 position can have a good contribution to allergenicity (yes) in the six allergens except for P04776. Especially, for the allergen P26987, the hydrophobicity, bulk, and electronic property of amino acid at position p6 promoted the allergenicity (yes).

In the soybean allergens including P04347, P11827, P25974, P26987, the most important amino acid property for allergenicity (yes) is z1 (hydrophobicity), followed by z2 (bulk) and z3 (electronic property). Electronic property is the most important amino acid property for the allergen P05046, whereas bulk is the most important amino acid property for the allergens of P01070 and P04776.

Except for the Y- variable allergenicity (yes), the random forest models also provided the relationship between X- variable and Y- variable allergenicity (total) to find important variables. From the Fig. 2, we found that the variable p1z1 can affect both allergenicity (no) and allergenicity (yes) in most allergens (P01070, P04776, P05406, and P25974), and the variable p6z1 contributed to both allergenicity (no) and allergenicity (yes) in three allergens (P04347, P11827, and P26987).

3.4 Simulated pepsin digestion of T cell epitopes

Most of the soybean proteins would be broken down by pepsin in the stomach, and only the anti-digestive peptides containing epitopes could enter the gastric mucosa and interact with immune cells to trigger allergic reactions [35]. Thus, the Peptidecutter tool was used to simulate the gastric digestion of seven soybean allergens, respectively, and the pepsin hydrolysis sites of the T cell epitopes are shown in Table 9, which further confirm the potential allergenicity.

As shown in Table 9, except for the T cell epitope "SKDNVISQIPSQVQE" in P11827 protein, all other epitopes contained multiple pepsin hydrolysis sites, and most of the digested peptides lengths were less than 12 amino acids. Since the length of the T cell epitope is generally believed between 12-20 amino acids [15], the peptide sequences of "DDGTRRLVVSKNKP" (aa 179-192) in P01070 protein, "TPVVAVSIIDTNS" (aa 158-170), "RHNIGQTSSPDI" (aa 322-333), and SQQARQIKNNNP" (aa 468-479) in P04776 protein, "NPHIGINVNSIRSIKTTS" (aa 168-185) in P05046 protein, "SKDNVISQIPSQVQE" (aa 558-572) in P11827 protein, "EEQRQQEGVIVE" (aa 201-212) in P25974 protein and the "LVTDADNVIPKA" (aa 23-34) in P26987 protein would be the digestion resistant regions. Among the 8 anti-digestive regions, isoleucine (I), valine (V), serine (S), asparagine (N), and glutamine (Q) were the amino acids with the highest content, while isoleucine (I), valine (V), and asparagine (N) covered 7 anti-digestive areas and serine (S) covered 6 anti-digestive areas. Interestingly, all 8 anti-digestive regions are exposed on the surface of the proteins except "TPVVAVSIIDTNS" (aa 158-170) in P04776 (shown in Tables 2-8), and as shown in Table 9, six peptides remain allergenicity among these 8 peptides. It has been reported that amino acids which had the high surface exposure were likely to bind with antibodies and formed B cell epitopes, however, the relationship between T cell epitopes and surface exposure has not been elaborated yet.

4. Discussion

At present, the food allergy is not exactly explained, which has hindered the development of hypoallergenic foods and immunotherapy [36, 37]. The identification of T cell epitopes can provide important information about the cellular mechanisms involved in the transition from food tolerance to allergy [38], and the analysis of cytokines related to Th2 cells can also improve our understanding of the immune response. Soybean

is one of the most widely cultivated edible legumes. In the past 50 years from 1968 to 2018, the world soybean production increased 8.4 times, and the global planted area also increased 4.3 times, while global soybean consumption increased from 24 million tons to 57 million tons between 1999 and 2019 [39]. Soybean is rich in carbohydrates, protein, fat, cellulose, and trace elements such as calcium and magnesium [40]. Although soybean is the high-quality protein source for humans and animals, the prevalence of sensitization of the major allergens from soybean has aroused wide attention. After entering the human body, soybean protein allergens will be treated by lysosomes and then combined with MHC class II molecules, and then the MHC class II-T cell epitope complex will interact with CD4+ T cells to synthesize and secrete the cytokine such as IL4, which will play a key role in promoting the proliferation and differentiation of antigen-presenting cells and stimulating the production of IgE [22]. In this study, the T cell epitopes of seven soybean protein allergens were comprehensively identified based on the binding affinity between epitopes and HLA II molecular, the inducing ability of IL4 production, and the antigenicity analysis, while the larger fragments released by pepsin digestion were also analyzed. As far as we have known, it might be the first paper t to analyze T cell epitopes of soybean allergens.

The combination of consensus method and NetMHCIIpan in IEDB have been used to predict T cell epitopes of allergens that can bind to MHC II molecules, and the NetMHCIIpan method has been recognized as the most accurate epitope prediction method [41]. Ramesh et al. [14] predicted 54 T cell epitopes of peanut allergen Ara h 1 by NetMHCIIpan method and identified 32 T cell epitopes were confirmed by using ProImmune REVEAL assay, indicated the higher accuracy (32 confirmed/37 predicted) of the NetMHCIIpan method. Pascal et al. [42] identified 4 T cell epitope regions of peanut allergen Ara h 2 via T cell proliferation assay and cytokine profile analysis, and all of the T cell epitopes were further proved to be MHC II binders by NetMHCIIpan method and NetMHCII method. In order to further improve accuracy on the prediction of T cell epitopes, combining consensus method and the NetMHCIIpan method in the IEDB database with other bioinformatics tools were explored in the previous studies, and all the predicted 6 T cell epitopes exhibited good potentials in T-cell proliferation and cytokines release [43]. Except for previous T cell epitopes prediction methods, the allergenicity and the ability to promote Th2 response-related cytokines IL-4 release of T cell epitopes were also evaluated in the current studies, which would further improve the prediction accuracy.

As shown in Tables 2-8, since the literature on T cell epitopes mapping of soybean has not been available until now, all the T-cell epitopes are first identified. The soybean allergen Gly m 4 belongs to the PR family and has cross-reactivity with the birch pollen allergen Bet v 1. Jahn- Schmid et al. [44] found that the "TLLRAVESYLLAHSD" (aa 142-156) of the Bet v 1 allergen was the major T cell epitope and could cross-react with the peptide segment of soybean Gly m 4 "ALFKAIEAYLLAHPD" (aa 142-156), and this peptide fragment is consistent with the T cell epitopes "ADALFKAIEAYLLAH" (aa 140-154) and "DALFKAIEAYLLAHP" (aa 141-155) of Gly m 4 predicted in this study. In addition, comparing with the conformational epitope of Gly m 4 reported by Husslik et al. [45], many conformational epitope amino acid residues were exited in the Gly m 4 T cell epitope regions "LYKALVTDADNVIPKA" (aa 19-34), "KKITFLEDGETKFVLHKIESI" (aa 54-74), "AGPNGGSAGKLTVKY" (aa 106-120), and "AKADALFKAIEAYLL" (aa 138-152) reported in the current experiment.

Allergen-specific T cells play an important role in allergic reactions and are obvious targets for immunotherapy intervention in diseases [17]. Recently, Candreva et al. [46] developed a new strategy for preventing milk allergy on the basis of oral administration of a soybean-derived peptide that was cross-reactive with bovine caseins. This peptide contained both T and B cell epitopes of soybean allergen Gly-m-Bd-30K and could stimulate T cells without causing IgE cross-linking on basophils and mast cells. In this study, the T cell epitope region "GGSILSGFTLEFLEHAFSVD" (217-236) in the P04776 allergen reported in this experiment is consistent with the previous results of the B cell epitope "GGSILSGFTLEFLEHAFSV" (217-235) identified by peptide scanning [47]. The T cell epitope regions "EEQRQQEGVIVELSK" (aa 201-215), "NPIYSNNFGKFFEIT" (aa 246-260), and "DIFLSSVDINEGALLLPHFNS" (aa 271-291) in P25974 allergen reported in this experiment are consistent with the epitopes identified by co-immunoprecipitation and mass spectrometry [48]. These T cell epitopes which contained B cell epitope amino acids would be expected for allergy

treatment via Synthetic Peptide Immuno-Regulatory Epitopes for related cross-reactive soybean.

Selecting potential T cell epitopes would be a difficult task because of the complexity of HLA alleles. In this study, since the epitopes "YIKDVFRVIPSEVLS" (aa 477-491), "KDVFRVIPSEVLSNS" (aa 479-493), "DVFRVIPSEVLSNSY" (aa 480-494) in P04347 protein could bind with more than 9 HLA molecules, and the epitopes "AKADALFKAIEAYLL" (aa 138-152), "ADALFKAIEAYLLAH" (aa 140-154) in P26987 protein could bind with more than 13 HLA molecules, and these epitopes can be predicted by all three methods (shown in Tables 3 and 8), the five new fragments are considered as the most possible epitope candidates.

Compared with the entire protein sequence, the frequency of phenylalanine (F), isoleucine (I), asparagine (N), valine (V) lysine (K), and histidine (H) increases in T cell epitopes of most soybean allergens. According to previous studies, the presence of lysine (K) had a significant effect on T cell stimulation and secondary structure alpha-helix could promote the antigenicity [49]. Another study also found that isoleucine (I) and histidine (H) was the key amino acids in the T cell epitope of the dust mite allergen Der p 2, because they might direct contact moieties, and might also indirectly affect peptide binding by changing the conformation of adjacent amino acid side chains [50].

A previous study showed that MHC molecular tended to bind hydrophobic amino acids in most positions except penultimate position, and could bind both hydrophilic and hydrophobic amino acids in positions 4 and 6, which could bind hydrophobic amino acids in positions 1, 2, 3, 5, and 8 [7]. Similarly, in the present paper, random forest models also showed that amino acids in positions p1, p2, p4, p5, p6, and p13 had a good contribution to the allergenicity. For positions p4, p5, and p6, three physicochemical properties (z1, z2, and z3) contributed to allergenicity in a similar way. For positions p1 and p2, the physicochemical properties that contribute to allergenicity were in the order z1>z2>z3, whereas the order for position p13 was z3>z2>z1. Furthermore, the hydrophobic (z1) residues at positions p1 and p2 contributed to the allergenicity in most soybean allergen models. In the position p1 of P04776, P05046, P11827, and P25974 models and the position p2 of P04347, P04776, P11827, and P26987, there are many hydrophobic residues including which can promote allergenicity (shown in Tables 2-8 and Fig. 2). Positively charged amino acids, such as arginine (R), lysine (K), and histidine (H) tend to locate at the position p13 and play an important role in the allergenicity in the soybean models P04347, P04776, P25974, and P05046. In this study, the frequency of phenylalanine (F), isoleucine (I), asparagine (N), valine (V) lysine (K), and histidine (H) residues increased in the T cell epitope region (Fig. 1), which was consistent with the current analysis, as hydrophobic residues including phenylalanine (F), isoleucine (I), and valine (V) located at the positions p1 and p2 contributed to the allergenicity, whereas positively charged amino acids, such as lysine (K), and histidine (H) located at the positions p6 and p13 and promoted the allergenicity (shown in Tables 3-5, Table 7, and Fig. 2), and previous studies also confirmed the positively charged amino acids would provide a net charge to increase the activity of the epitope peptide antigen [50].

In simulated pepsin digestion experiments, most of the T cell epitopes from soybean allergen could be hydrolyzed by pepsin into small peptides (<12 aa), and most of the anti-digestive fragments are located on the surface of the proteins. The digestion-resistant epitope region contained much hydrophobic amino acids including isoleucine (I), valine (V), and non-charged amino acids including serine (S), asparagine (N), and glutamine (Q), and the second position of this peptide with hydrophobic amino acid might contribute to the allergenicity, such as the anti-digestive fragment "LVTDADNVIPKA" (aa 23-34) in P26987. The T cell epitopes EEQRQQEGVIVELSK" (aa 201-215), and "EQRQQEGVIVELSKE" (aa 202-216) with higher IL-4pred scores also can be resistant to pepsin hydrolysis and have a great potential to enter into bodies to cause Th2 cell response. In order to development of soybean hypoallergenic products, proteases that can hydrolyze the above five amino acid sites (I, V, S, N, and Q) can be selected. For example, the endopeptidase Glu-C can hydrolyze glutamate (E) and glutamine (Q) [51, 52, 53] can be used to specifically destroy the digestion-resistant T cell epitopes "SKDNVISQIPSQVQE" (aa 558-572) (P11827) and EEQRQQEGVIVE" (aa 201-212) (P25974).

5. Conclusions

T cell epitopes of seven soybean allergens were predicted by using the"MHC-II Binding Predictions" tools in the IEDB database, and 138 T cell epitopes were obtained by the further evaluation of the ability to induce the production of IL4 and allergenicity by bioinformatics tools, and 5 predicted fragments of "YIKDVFRVIPSEVLS" (aa 477-491), "KDVFRVIPSEVLSNS" (aa 479-493), "DVFRVIPSEVLSNSY" (aa 480-494) in P04347 protein and "AKADALFKAIEAYLL" (aa 138-152), "ADALFKAIEAYLLAH" (aa 140-154) in P26987 protein were considered as the most possible epitope candidates. Based on the amino acid composition analysis and random forest model, hydrophobic residues (such as phenylalanine, isoleucine, and valine) at positions p1 and p2 and positively charged amino acid residues (such as lysine and histidine) in positions p13 would have a good contribution to the allergenicity. Furthermore, most of the epitopes of T cells could be hydrolyzed by pepsin into small molecular peptides (<12 aa), and anti-digestive epitope regions contained more isoleucine (I), valine (V), serine (S), asparagine (N), and glutamine (Q). In the current studies, bioinformatics strategies provide an efficient method for epitope prediction and analysis, and the results may aid to study the mechanism of soybean sensitization. Additionally, epitopes reported here are expected to apply in immunotherapy designing and hypoallergenic soybean products developing in the food industry.

Abbreviation

WHO/IUIS, World Health Organization and the International Union of Immunological Societie; MHC, Major Histocompatibility Complexs; HLA, Human Leukocyte Antigens; IEDB, Immune Epitope Database Analysis Resource; ANN, Artificial Neural Network; SMM, Stabilization Matrix

Access dates of websites

Allergen nomenclature. WHO/IUIS Allergen Nomenclature Sub-Committee. (2020). http://www.allergen.org/ Accessed 21 July 2020. Uniprot database. (2020). https://www.uniprot.org/ Accessed 10 September 2020. IL4pred tool. (2020). https://webs.iiitd.edu.in/raghava/il4pred/ Accessed 23 November 2020. AllerTop 2.0. (2020). https://www.ddg-pharmfac.net/AllerTOP/ Accessed 2 November 2020. PeptideCutter. (2020). https://web.expasy.org/peptide_cutter/ Accessed 25 December 2020.

Conflicts of interest

The authors declare no conflict of interest.

Acknowledgments

The authors thank the hard work and diligence of the predecessors who worked in the food protein allergen area and appreciate the helpful comments from the anonymous reviewers. The authors are also grateful for the financial support from the National Natural Science Foundation of China (No. 31701524), Major Special Science and Technology Planning Project of Anhui Province, China (No. 202003b06020030), Key Research and Development Program of Anhui Province, China (No. 202104a06020016).

References

[1] S.H. Sicherer, H.A. Sampson, Food Allergy: A review and update on epidemiology, pathogenesis, diagnosis, prevention and management, Journal of Allergy & Clinical Immunology, 141 (2017) 41-58. doi: 10.1016/j.jaci.2017.11.003.

[2] F. Zhou, S. He, H. Sun, Y. Wang, Y. Zhang, Advances in epitope mapping technologies for food protein allergens: A review, Trends in Food Science & Technology, 107 (2020) 226-239. doi: 10.1016/j.tifs.2020.10.035.

[3] X. Sun, X. Shan, Z. Yan, Y. Zhang, L. Guan, Prediction and characterization of the linear IgE epitopes for the major soybean allergen β-conglycinin using immunoinformatics tools, Food & Chemical Toxicology, 56 (2013) 254-260. doi: 10.1016/j.fct.2013.02.014.

[4] G. Bu, T. Li, T. Zhu, G. Xi, Identification of the linear immunodominant epitopes in the β subunit of β-conglycinin and preparation of epitope antibodies, International Journal of Biological Macromolecules, 154 (2020) 724-731. doi: 10.1016/j.ijbiomac.2020.03.159.

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

[6] Malherbe, Laurent, T-cell epitope mapping, Annals of Allergy Asthma & Immunology, 103 (2009) 76-79. doi: 10.1016/S1081-1206(10)60147-0

[7] P. Walden, T-cell epitope determination, Current Opinion in Immunology, 8 (1996) 68. doi: 10.1016/S0952-7915(96)80107-5.

[8] W.-L. Fan, M.-S. Shiao, R.C.-Y. Hui, S.-C. Su, C.-W. Wang, Y.-C. Chang, W.-H. Chung, HLA association with drug-induced adverse reactions, Journal of Immunology Research, 2017 (2017) 1-10. doi: 10.1155/2017/3186328.

[9] T. Rahaman, T. Vasiljevic, L. Ramchandran, Effect of processing on conformational changes of food proteins related to allergenicity, Trends in Food ence & Technology, 49 (2016) 24-34. doi: 10.1016/j.tifs.2016.01.001.

[10] G. Bu, Y. Luo, F. Chen, K. Liu, T. Zhu, Milk processing as a tool to reduce cow's milk allergenicity: a mini-review, Dairy Science & Technology, 93 (2013) 211-223. doi: 10.1007/s13594-013-0113-x.

[11] C.Y.Y. Wai, N.Y.H. Leung, P.S.C. Leung, K.H. Chu, T cell epitope immunotherapy ameliorates allergic responses in a murine model of shrimp allergy, Clinical & Experimental Allergy, 46 (2016) 491-503. doi: 10.1111/cea.12684.

[12] P. Rupa, Y. Mine, Oral immunotherapy with immunodominant T-cell epitope peptides alleviates allergic reactions in a Balb/c mouse model of egg allergy, Allergy, 67 (2012) 74-82. doi: 10.1111/j.1398-9995.2011.02724.x.

[13] R.E. Soria-Guerra, R. Nieto-Gomez, D.O. Govea-Alonso, S. Rosales-Mendoza, An overview of bioinformatics tools for epitope prediction: Implications on vaccine development, Journal of Biomedical Informatics, 53 (2015) 405-414. doi: 10.1016/j.jbi.2014.11.003.

[14] M. Ramesh, A. Yuenyongviwat, G.N. Konstantinou, J. Lieberman, M. Pascal, M. Masilamani, H.A. Sampson, Peanut T-cell epitope discovery: Ara h 1, Journal of Allergy and Clinical Immunology, 137 (2016) 1764-1771. doi: 10.1016/j.jaci.2015.12.1327.

[15] T.A. Ahmad, A.E. Eweida, L.H. El-Sayed, T-cell epitope mapping for the design of powerful vaccines, Vaccine Reports, 6 (2016) 13-22. doi: 10.1016/j.vacrep.2016.07.002.

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

[17] S. He, J. Zhao, W. Elfalleh, M. Jemaà, H. Sun, X. Sun, M. Tang, Q. He, Z. Wu, F. Lang, In silico identification and in vitro analysis of B and T-cell epitopes of the black turtle bean (Phaseolus vulgaris L.) lectin, Cellular Physiology & Biochemistry, 49 (2018) 1600-1614. doi: 10.1159/000493496.

[18] M.G. Zeece, T.A. Beardslee, J.P. Markwell, G. Sarath, Identification of an IgE-binding region in soybean acidic glycinin G1, Food & Agricultural Immunology, 11 (1999) 83-90. doi: 10.1080/09540109999942.

[19] R.M. Helm, G. Cockrell, C. Connaughton, H.A. Sampson, G.A. Bannon, V. Beilinson, N.C. Nielsen, A.W. Burks, A soybean G2 glycinin allergen, International Archives of Allergy and Immunology, 123 (2000) 205-212. doi: 10.1159/000024446.

[20] H. Saeed, C. Gagnon, E. Cober, S. Gleddie, Using patient serum to epitope map soybean glycinins reveals common epitopes shared with many legumes and tree nuts, Molecular Immunology, 70 (2016) 125-133. doi: 10.1016/j.molimm.2015.12.008.

[21] R. Vita, J.A. Overton, J.A. Greenbaum, J. Ponomarenko, J.D. Clark, J.R. Cantrell, D.K. Wheeler, J.L. Gabbard, D. Hix, A. Sette, B. Peters, The immune epitope database (IEDB) 3.0, Nucleic Acids Research,

43 (2015) 405-412. doi: 10.1093/nar/gku938.

[22] S.K. Dhanda, S. Gupta, P. Vir, G.P.S. Raghava, Prediction of IL4 inducing peptides, Clinical & Developmental Immunology, 2013 (2013) 1-9. doi: 10.1155/2013/263952.

[23] I. Dimitrov, D.R. Flower, I. Doytchinova, AllerTOP v.2–a server for in silico prediction of allergens, BMC Bioinformatics, 20 (2014) 1-6. doi: 10.1186/1471-2105-14-S6-S4.

[24] K. Narayanan, In silico prediction of potential vaccine candidates on capsid protein of human bocavirus 1, Molecular Immunology, 93 (2018) 193-205. doi: 10.1016/j.molimm.2017.11.024.

[25] L.-N. Zheng, H. Lin, R. Pawar, Z.-X. Li, M.-H. Li, Mapping IgE binding epitopes of major shrimp (Penaeus monodon) allergen with immunoinformatics tools, Food & Chemical Toxicology, 49 (2011) 2954-2960. doi: 10.1016/j.fct.2011.07.043.

[26] S. Hellberg, M. Sjoestroem, B. Skagerberg, S. Wold, Peptide quantitative structure-activity relationships, a multivariate approach, Journal of Medicinal Chemistry, 30 (1987) 1126-1135. doi: 10.1021/jm00390a003.

[27] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M.R. Wilkins, R.D. Appel, A. Bairoch, Protein identification and analysis tools on the ExPASy server, (In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press (2005) pp. 571-607.

[28] Sinu Paul, C.S.L. Arlehamn, T.J. Scriba, M.B.C. Dillon, C. Oseroff, D. Hinz, D.M. McKinney, S.C. Pro, J. Sidney, B. Peters, A. Sette, Development and validation of a broad scheme for prediction of HLA class II restricted T cell epitopes, Journal of Immunological Methods, 422 (2015) 28-34. doi: 10.1016/j.jim.2015.03.022.

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

[30] S. Natarajan, C. Xu, H. Bae, B.A. Bailey, P. Cregan, T.J. Caperna, W.M. Garrett, D. Luthria, Proteomic and genetic analysis of glycinin subunits of sixteen soybean genotypes, Plant Physiology & Biochemistry, 45 (2007) 436-444. doi: 10.1016/j.plaphy.2007.03.031.

[31] T. Wang, G.-x. Qin, Z.-w. Sun, Y. Zhao, Advances of research on glycinin and β-conglycinin: A review of two major soybean allergenic proteins, Critical Reviews in Food Science & Nutrition, 54 (2014) 850-862. doi: 10.1080/10408398.2011.613534.

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

[33] H.B. KRISHNAN, W.-S. KIM, S. JANG, M.S. KERLEY, All three subunits of soybean β-conglycinin are potential food allergens, Journal of Agricultural and Food Chemistry, 57 (2009) 938-943. doi: 10.1021/jf802451g.

[34] T. Geng, K. Liu, R. Frazier, L. Shi, E. Bell, K. Glenn, J.M. Ward, Development of a sandwich ELISA for quantification of Gly m 4, a soybean allergen, Journal of Agricultural & Food Chemistry, 63 (2015) 4947-4953. doi: 10.1021/acs.jafc.5b00792.

[35] J. Huang, C. Liu, Y. Wang, C. Wang, M. Xie, Y. Qian, L. Fu, Application of in vitro and in vivo models in the study of food allergy, Food Science & Human Wellness, 7 (2018) 235-243. doi: 10.1016/j.fshw.2018.10.002.

[36] L.T. Villuendas, J. Cuestaherranz, M. Gonzalezmuñoz, L.F. Pacios, E. Compes, B. Garciacarrasco, G.S. Duran, A.D. Perales, T-cell epitopes of the major peach allergen, Pru p 3: Identification and differential T-cell response of peach-allergic and non-allergic subjects, Molecular Immunology, 46 (2009) 722-728. doi: 10.1016/j.molimm.2008.10.018.

[37] Characterization of the T-cell epitopes of a major peanut allergen, Ara h 2, Allergy, 60 (2005) 35-40. doi: 10.1111/j.1398-9995.2004.00608.x.

[38] E.A. Pastorello, M. Monza, V. Pravettoni, R. Longhi, P. Bonara, J. Scibilia, L. Primavesi, R. Scorza, Characterization of the T-cell epitopes of the major peach allergen Pru p 3, International Archives of Allergy and Immunology, 153 (2010) 1-12. doi: 10.1159/000301573.

[39] M. De Maria, E.J. Robinson, J.R. Kangile, R. Kadigi, I. Dreoni, M. Couto, N. Howai, J. Peci, Global soybean trade-the geopolitics of a bean, (2020). doi: 10.34892/7yn1-k494.

[40] N. Guzeler, C. Yildirim, The utilization and processing of soybean and soybean products, Journal of Agricultural Faculty of Uludag University, 30 (2016) 546-553.

[41] L. Zhang, K. Udaka, H. Mamitsuka, S. Zhu, Toward more accurate pan-specific MHC-peptide binding prediction: a review of current methods and tools, Briefings in Bioinformatics, 13 (2012) 350-364. doi: 10.1093/bib/bbr060.

[42] M. Pascal, G. Konstantinou, M. Masilamani, J. Lieberman, H. Sampson, In silico prediction of Ara h 2 T cell epitopes in peanut-allergic children, Clinical & Experimental Allergy, 43 (2013) 116-127. doi: 10.1111/cea.12014.

[43] L.L. Xu, M. Gasset, H. Lin, C. Yu, J.L. Zhao, X.W. Dang, Z.X. Li, Identification of the dominant T-cell epitopes of Lit v 1 shrimp major allergen and their functional overlap with known B-cell epitopes, Journal of Agricultural and Food Chemistry, 69 (2021) 7420–7428. doi: 10.1021/acs.jafc.1c02231.

[44] B. Jahn-Schmid, A. Radakovics, D. Luttkopf, S. Scheurer, S. Vieths, C. Ebner, B. Bohle, Bet v 1142-156 is the dominant T-cell epitope of the major birch pollen allergen and important for cross-reactivity with Bet v 1-related food allergens, Journal of Allergy and Clinical Immunology, 116 (2005) 213-219. doi: 10.1016/j.jaci.2005.04.019.

[45] F. Husslik, J. Nurnberg, C.S.v. Loetzen, T. Mews, B.K. BallmerWeber, J. Kleine-Tebbe, R. Treudler, J.-C. Simon, S. Randow, E. Volker, A. Reuter, P. Rosch, S. Vieths, T. Holzhauser, D. Schiller, The conformational IgE epitope profile of soya bean allergen Gly m 4, Clinical & Experimental Allergy, 46 (2016) 1484-1497. doi: 10.1111/cea.12796.

[46] N.M. Candreva, P.L. Smaldini, A. Cauerhff, S. Petruccelli, G.H. Docena, A novel approach to ameliorate experimental milk allergy based on the oral administration of a short soy cross-reactive peptide, Food Chemistry, 346 (2020) 128926. doi: 10.1016/j.foodchem.2020.128926.

[47] T.A. Beardslee, M.G. Zeece, G. Sarath, J.P. Markwell, Soybean glycinin G1 acidic chain shares IgE epitopes with peanut allergen Ara h 3, International Archives of Allergy & Immunology, 123 (2000) 299-307. doi: 10.1159/000053642.

[48] Y. Zhao, G. Naren, J. Qiang, G. Qin, N. Bao, M.H. Farouk, Identification of allergic epitopes of soybean β-conglycinin in different animal species, Frontiers in Veterinary Science, 7 (2020) 599546. doi: 10.3389/fvets.2020.599546.

[49] I. Berkower, G.K. Buckenmeyer, J.A. Berzofsky, Molecular mapping of a histocompatibility-restricted immunodominant T cell epitope with synthetic and natural peptides: implications for T cell antigenic structure, Journal of Immunology, 136 (1986) 2498-2503. doi: 10.1084/jem.163.4.1030.

[50] B. Wu, G. Toussaint, L.V. Elst, C. Granier, M.G. Jacquemin, J.-M.R. Saint-Remy, Major T cell epitope-containing peptides can elicit strong antibody responses, European Journal of Immunology, 30 (2000) 291-299. doi: 10.1002/1521-4141(200001)30:1<291::AID-IMMU291>3.0.CO;2-T.

[51] M. Weiss, B. Dhayalan, Y.-S. Chen, N. Phillips, M. Swain, N. Rege, A. Mirsalehi, M. Jarosinski, F. Ismail-Beigi, N. Metanis, Reassessment of an innovative insulin analogue excludes protracted action yet

highlights the distinction between external and internal diselenide bridges, Chemistry-A European Journal, 26 (2020) 4695. doi: 10.1023/A:1007042825783.

[52] N. Priolo, S.M.d. Valle, M.C. Arribére, L. López, a.N. Caffini, Isolation and characterization of a cysteine protease from the latex of Araujia hortorum fruits, Journal of Protein Chemistry, 19 (2000) 39-49. doi: 10.1023/A:1007042825783.

[53] C.E. Osorio, N. Wen, J.H. Mejías, S. Mitchell, D.v. Wettstein, S. Rustgi, Directed-mutagenesis of flavobacterium meningosepticum prolyl-oligopeptidase and a glutamine-specific endopeptidase from barley, Frontiers in Nutrition, 7 (2020) 11. doi: 10.3389/fnut.2020.00011.

Table/Figure Captions

Table 1 The names, structures, and related properties of seven soybean allergens.

Table 2 List of putative epitopes of P01070 along with interacting HLA class II alleles and relevant scores.

Table 3 List of putative epitopes of P04347 along with interacting HLA class II alleles and relevant scores.

Table 4 List of putative epitopes of P04776 along with interacting HLA class II alleles and relevant scores.

Table 5 List of putative epitopes of P05046 along with interacting HLA class II alleles and relevant scores.

Table 6 List of putative epitopes of P11827 along with interacting HLA class II alleles and relevant scores.

Table 7 List of putative epitopes of P25974 along with interacting HLA class II alleles and relevant scores.

Table 8 List of putative epitopes of P26987 along with interacting HLA class II alleles and relevant scores.

Table 9 Pepsin cleavage site of T cell epitopes.

Fig. 1 Analysis of the amino acid composition of T cell epitopes of seven soybean allergens.

(P01070: Amino acid composition analysis of P01070 protein; P01070-T: Amino acid composition analysis of T cell epitopes of P01070 protein. P04347: Amino acid composition analysis of P04347 protein; P04347-T: Amino acid composition analysis of T cell epitopes of P04347 protein. P04776: Amino acid composition analysis of P04776 protein; P04776-T: Amino acid composition analysis of T cell epitopes of P04776 protein. P05046: Amino acid composition analysis of P05046 protein; P05046-T: Amino acid composition analysis of T cell epitopes of P05046 protein. P11827: Amino acid composition analysis of P11827 protein; P11827-T: Amino acid composition analysis of T cell epitopes of P11827 protein. P25974: Amino acid composition analysis of P25974 protein; P25974-T: Amino acid composition analysis of T cell epitopes of P P25974 protein. P26987: Amino acid composition analysis of P26987 protein; P26987-T: Amino acid composition analysis of T cell epitopes of P26987 protein.)

Fig. 2 Mean decrease accuracy of the 3-z scale models of the (A) P01070 T cell epitopes, (B) P04347 T cell epitopes, (C) P04776 T cell epitopes, (D) P05046 T cell epitopes, (E) P11827 T cell epitopes, (F) P25974 T cell epitopes, and (G) P26987 T cell epitopes.

Table 1 The names, structures, and related properties of seven soybean allergens.

Protein name	**IUIS name**	**Uniprot name**	**Number of amino acids**	**Molecular w**
Pathogenesis-related protein	Gly m 4.0101	P26987	158	16772
β-conglycinin	Gly m 5.0201	P11827	481	54242
	Gly m 5.0301/Gly m 5.0302	P25974	439	50476
Glycinin	Gly m 6.0101	P04776	495	55706
	Gly m 6.0501	P04347	516	57956
Kunitz trypsin inhibitor	-	P01070	216	24005
Lectin	-	P05046	285	30928

Table 2 List of putative epitopes of P01070 along with interacting HLA class II alleles and relevant scores.

HLA class II alleles
HLA-DRB1*01:01/ HLA-DQA1*05:01/ HLA-DQB1*03:01
HLA-DQA1*05:01/ HLA-DQB1*03:01
HLA-DRB1*01:01/ HLA-DQA1*05:01/ HLA-DQB1*03:01
HLA-DRB3*02:02
HLA-DRB3*02:02
HLA-DPA1*01:03/ HLA-DPB1*04:01/ HLA-DPA1*01:03/ HLA-DPB1*02:01
HLA-DPA1*01:03/ HLA-DPB1*02:01/ HLA-DPA1*03:01/ HLA-DPB1*04:02/ HLA-DPA1*02:01/ HLA-DPB1*01:01/ HLA
HLA-DPA1*01:03/ HLA-DPB1*02:01/ HLA-DPA1*03:01/ HLA-DPB1*04:02/ HLA-DPA1*02:01/ HLA-DPB1*01:01/ HLA
HLA-DRB1*07:01/HLA-DRB1*13:02
HLA-DRB1*13:02/ HLA-DRB1*07:01
HLA-DRB1*13:02/ HLA-DRB1*07:01/ HLA-DRB1*15:01
HLA-DRB1*13:02/ HLA-DRB1*07:01/ HLA-DRB1*15:01
HLA-DRB1*13:02/ HLA-DRB1*07:01/ HLA-DRB3*02:02
HLA-DRB1*13:02/ HLA-DRB3*02:02

Note: A: Combination method of SMM method ANN method, and combinatorial library method; B: Combination method of SMM method, ANN method, and Sturniolo method; C: NetMHCIIpan method.

Table 3 List of putative epitopes of P04347 along with interacting HLA class II alleles and relevant scores.

HLA class II alleles
HLA-DRB3*02:02
HLA-DRB3*02:02
HLA-DRB3*02:02
HLA-DPA1*01:03/ HLA-DPB1*04:01
HLA-DPA1*01:03/ HLA-DPB1*04:01
HLA-DQA1*05:01/ HLA-DQB1*03:01
HLA-DQA1*05:01/ HLA-DQB1*03:01
HLA-DRB1*15:01/ HLA-DRB1*11:01
HLA-DRB1*15:01/ HLA-DRB1*11:01
HLA-DRB1*15:01/ HLA-DRB1*11:01
HLA-DRB1*15:01/ HLA-DRB1*11:01
HLA-DRB1*04:01/ HLA-DRB3*02:02
HLA-DRB3*02:02/ HLA-DRB1*04:01
HLA-DRB1*04:01/ HLA-DRB3*02:02
HLA-DRB3*02:02/ HLA-DRB1*04:01
HLA-DRB3*01:01
HLA-DRB3*01:01
HLA-DRB3*01:01
HLA-DRB1*09:01/ HLA-DRB1*07:01/ HLA-DRB1*15:01/ HLA-DRB5*01:01/ HLA-DPA1*02:01/ HLA-DPB1*14:01/ HLA
HLA-DRB1*09:01/ HLA-DRB1*07:01/ HLA-DRB1*15:01/ HLA-DRB1*01:01/ HLA-DRB1*04:01/ HLA-DRB1*04:05/ HLA
HLA-DRB1*09:01/ HLA-DRB1*07:01/ HLA-DRB1*15:01/ HLA-DRB1*04:01/ HLA-DRB5*01:01/HLA-DPA1*02:01/ HLA
HLA-DPA1*02:01/ HLA-DPB1*14:01/ HLA-DRB3*02:02

Note: A: Combination method of SMM method ANN method, and combinatorial library method; B: Combination method of SMM method, ANN method, and Sturniolo method; C: NetMHCIIpan method.

Table 4 List of putative epitopes of P04776 along with interacting HLA class II alleles and relevant scores.

HLA class II alleles
HLA-DRB4*01:01
HLA-DPA1*02:01/ HLA-DPB1*01:01/ HLA-DPA1*01:03/ HLA-DPB1*02:01/ HLA-DPA1*03:01/ HLA-DPB1*04:02/ HLA
HLA-DPA1*02:01/ HLA-DPB1*01:01/ HLA-DPA1*03:01/ HLA-DPB1*04:02/ HLA-DPA1*01:03/ HLA-DPB1*02:01/ HLA
HLA-DPA1*02:01/ HLA-DPB1*01:01/ HLA-DPA1*03:01/ HLA-DPB1*04:02/ HLA-DPA1*01:03/ HLA-DPB1*02:01/ HLA
HLA-DPA1*03:01/ HLA-DPB1*04:02/ HLA-DPA1*01:03/ HLA-DPB1*02:01/ HLA-DPA1*01:03/ HLA-DPB1*04:01
HLA-DRB3*02:02
HLA-DPA1*03:01/ HLA-DPB1*04:02/ HLA-DPA1*01:03/ HLA-DPB1*04:01
HLA-DPA1*03:01/ HLA-DPB1*04:02
HLA-DRB3*02:02
HLA-DRB1*04:01/ HLA-DRB1*13:02/ HLA-DRB1*07:01/ HLA-DRB3*02:02
HLA-DRB1*13:02/ HLA-DRB3*01:01/ HLA-DRB1*07:01/ HLA-DRB3*02:02
HLA-DRB1*13:02/ HLA-DRB3*01:01/ HLA-DRB3*02:02
HLA-DQA1*01:02/ HLA-DQB1*06:02/ HLA-DRB1*13:02/ HLA-DRB3*01:01/ HLA-DRB3*02:02
HLA-DQA1*01:02/ HLA-DQB1*06:02/ HLA-DRB1*13:02/ HLA-DRB3*01:01/ HLA-DRB3*02:02
HLA-DQA1*01:02/ HLA-DQB1*06:02/ HLA-DRB1*13:02/ HLA-DRB3*02:02
HLA-DRB1*11:01
HLA-DRB1*11:01
HLA-DRB1*11:01
HLA-DRB1*11:01
HLA-DRB1*11:01
HLA-DRB1*11:01
HLA-DPA1*01:03/ HLA-DPB1*02:01/ HLA-DPA1*02:01/ HLA-DPB1*01:01/ HLA-DPA1*01:03/ HLA-DPB1*04:01
HLA-DPA1*01:03/ HLA-DPB1*02:01/ HLA-DPA1*02:01/ HLA-DPB1*01:01
HLA-DRB1*04:05/ HLA-DPA1*01:03/ HLA-DPB1*02:01/ HLA-DPA1*02:01/ HLA-DPB1*01:01
HLA-DRB1*04:05/ HLA-DPA1*01:03/ HLA-DPB1*02:01/ HLA-DPA1*02:01/ HLA-DPB1*01:01
HLA-DRB1*04:05/ HLA-DRB1*04:01/ HLA-DRB3*01:01/ HLA-DRB1*13:02/ HLA-DRB3*02:02
HLA-DRB1*13:02
HLA-DRB1*13:02/ HLA-DRB3*01:01/ HLA-DRB3*02:02
HLA-DRB1*13:02/ HLA-DRB3*01:01/ HLA-DRB3*02:02
HLA-DRB1*13:02/ HLA-DRB3*01:01/ HLA-DRB3*02:02
HLA-DRB1*13:02/ HLA-DRB3*02:02
HLA-DRB1*04:05/ HLA-DRB1*13:02
HLA-DRB1*04:05
HLA-DRB1*04:05

Note: A: Combination method of SMM method ANN method, and combinatorial library method; B: Combination method of SMM method, ANN method, and Sturniolo method; C: NetMHCIIpan method.

Table 5 List of putative epitopes of P05046 along with interacting HLA class II alleles and relevant scores.

HLA class II alleles	**Method**	**Position**
HLA-DRB3*02:02	C	23-37
HLA-DPA1*01:03/ HLA-DPB1*02:01/ HLA-DPA1*01:03/ HLA-DPB1*04:01	A,C	34-48
HLA-DRB3*02:02	C	38-52
HLA-DRB3*02:02	C	40-54
HLA-DRB1*13:02	B	78-92
HLA-DQA1*05:01/ HLA-DQB1*03:01	A	92-106

HLA class II alleles	Method	Position
HLA-DQA1*05:01/ HLA-DQB1*03:01	A	93-107
HLA-DQA1*05:01/ HLA-DQB1*03:01	A	94-108
HLA-DQA1*05:01/ HLA-DQB1*03:01	A	95-109
HLA-DPA1*01:03/ HLA-DPB1*02:01/ HLA-DRB1*04:05/ HLA-DPA1*01:03/ HLA-DPB1*04:01	A,B,C	100-114
HLA-DRB1*13:02/ HLA-DRB3*02:02	B,C	168-182
HLA-DRB1*13:02/ HLA-DRB3*02:02	B,C	169-183
HLA-DRB1*13:02/ HLA-DRB3*02:02	B,C	170-184
HLA-DRB1*13:02/ HLA-DRB3*02:02	B,C	171-185
HLA-DRB1*04:01/ HLA-DRB1*08:02/ HLA-DRB1*04:05	B	172-186
HLA-DRB1*04:01/ HLA-DRB1*04:05/ HLA-DRB1*08:02	B	173-187
HLA-DRB1*04:01/ HLA-DRB1*04:05	B	174-188
HLA-DRB1*13:02/ HLA-DRB3*02:02	B,C	184-198
HLA-DRB1*13:02/ HLA-DRB3*02:02	B,C	185-199
HLA-DRB1*13:02/ HLA-DRB3*02:02	B,C	186-200
HLA-DRB1*07:01/ HLA-DRB1*09:01/ HLA-DQA1*05:01/ HLA-DQB1*03:01	A	234-248
HLA-DRB1*07:01/ HLA-DRB1*09:01	A	236-250
HLA-DRB1*09:01	A	253-267
HLA-DRB1*09:01/ HLA-DRB3*02:02	A,C	254-268
HLA-DRB1*09:01/ HLA-DRB3*02:02	A,C	255-269

Note: A: Combination method of SMM method ANN method, and combinatorial library method; B: Combination method of SMM method, ANN method, and Sturniolo method; C: NetMHCIIpan method.

Table 6 List of putative epitopes of P11827 along with interacting HLA class II alleles and relevant scores.

HLA class II alleles	Method	Position
HLA-DPA1*03:01/ HLA-DPB1*04:02/ HLA-DRB1*11:01/ HLA-DPA1*01:03/ HLA-DPB1*04:01	A,B,C	211-225
HLA-DPA1*03:01/ HLA-DPB1*04:02/ HLA-DRB1*11:01/ HLA-DPA1*01:03/ HLA-DPB1*04:01	A,B,C	212-226
HLA-DPA1*01:03/ HLA-DPB1*04:01	C	213-227
HLA-DRB5*01:01	B	221-235
HLA-DRB5*01:01	B	222-236
HLA-DRB5*01:01	B	223-237
HLA-DRB5*01:01/ HLA-DRB1*11:01	B	225-239
HLA-DRB1*11:01/ HLA-DRB5*01:01	B	226-240
HLA-DRB1*04:01	B	248-262
HLA-DRB1*04:01/ HLA-DRB5*01:01/ HLA-DRB3*02:02	B,C	249-263
HLA-DRB3*01:01	A	262-276
HLA-DRB1*04:05	B	275-289
HLA-DRB1*04:05	B	276-290
HLA-DRB1*04:05	B	277-291
HLA-DQA1*05:01/ HLA-DQB1*03:01	A	520-534
HLA-DQA1*05:01/ HLA-DQB1*03:01	A	521-535
HLA-DRB4*01:01	A	558-572

Note: A: Combination method of SMM method ANN method, and combinatorial library method; B: Combination method of SMM method, ANN method, and Sturniolo method; C: NetMHCIIpan method.

Table 7 List of putative epitopes of P25974 along with interacting HLA class II alleles and relevant scores.

HLA class II alleles	Method	Position	Sequence	IL-4pred
HLA-DRB1*13:02/ HLA-DRB3*02:02	B,C	41-55	SFQTLFENQNGRIRL	0.43
HLA-DRB1*13:02/ HLA-DRB3*02:02	B,C	42-56	FQTLFENQNGRIRLL	0.22
HLA-DRB3*02:02	C	43-57	QTLFENQNGRIRLLQ	0.21
HLA-DRB1*15:01	B	71-85	DYRIVQFQSKPNTIL	0.34
HLA-DQA1*01:02/ HLA-DQB1*06:02	A	201-215	EEQRQQEGVIVELSK	1.31
HLA-DQA1*01:02/ HLA-DQB1*06:02	A	202-216	EQRQQEGVIVELSKE	1.32
HLA-DRB1*11:01	B	212-226	ELSKEQIRQLSRRAK	0.21
HLA-DRB3*01:01	A	244-258	SRNPIYSNNFGKFFE	0.42
HLA-DRB3*01:01	A	245-259	RNPIYSNNFGKFFEI	1.37
HLA-DRB3*01:01	A	246-260	NPIYSNNFGKFFEIT	0.41
HLA-DRB1*13:02/ HLA-DRB1*04:05	B	271-285	DIFLSSVDINEGALL	0.87
HLA-DRB1*13:02/ HLA-DQA1*01:02/ HLA-DQB1*06:02	A, B	275-289	SSVDINEGALLLPHF	0.27
HLA-DQA1*01:02/ HLA-DQB1*06:02	A	276-290	SVDINEGALLLPHFN	0.27
HLA-DQA1*01:02/ HLA-DQB1*06:02	A	277-291	VDINEGALLLPHFNS	0.27
HLA-DRB3*01:01	A	335-349	SEDDVFVIPAAYPFV	0.27

Note: A: Combination method of SMM method ANN method, and combinatorial library method; B: Combination method of SMM method, ANN method, and Sturniolo method; C: NetMHCIIpan method.

Table 8 List of putative epitopes of P26987 along with interacting HLA class II alleles and relevant scores.

HLA class II alleles
HLA-DRB3*01:01
HLA-DRB3*01:01/ HLA-DRB1*13:02
HLA-DPA1*02:01/ HLA-DPB1*01:01
HLA-DPA1*02:01/ HLA-DPB1*01:01
HLA-DRB3*01:01/ HLA-DRB1*03:01
HLA-DRB3*01:01/ HLA-DRB1*03:01
HLA-DRB3*01:01/ HLA-DRB1*03:01
HLA-DRB1*07:01
HLA-DQA1*05:01/ HLA-DQB1*03:01
HLA-DPA1*02:01/ HLA-DPB1*05:01/ HLA-DRB1*09:01/ HLA-DRB1*15:01/ HLA-DPA1*02:01/ HLA-DPB1*01:01/ HLA
HLA-DPA1*02:01/ HLA-DPB1*05:01/ HLA-DRB1*01:01/ HLA-DRB1*09:01/ HLA-DRB1*15:01/ HLA-DPA1*02:01/ HLA
HLA-DPA1*02:01/ HLA-DPB1*05:01

Note: A: Combination method of SMM method ANN method, and combinatorial library method; B: Combination method of SMM method, ANN method, and Sturniolo method; C: NetMHCIIpan method.

Table 9 Pepsin cleavage site of T cell epitopes.

Protein name	Position and sequence	Pepsin digestion sites	Anti-digestive peptides	Allergenicity of anti-digestive peptides
P01070	— — — 44-60 L SDITA F GGIRAAPTGN — — — — — — — — — — — 86-111 YRIRF IAEGHP LS L K F DS F AVIM L CV — — — — 179-199 DDGTR-RLVVSKNKP LVVQ F QK	44, 49, 50, 90, 96, 98, 99, 100, 101, 103, 104, 108, 109, 192, 196, 197	DDGTRRLVVSKNKP	Yes
P04347	— — — 183-199 NPRV FY L AGNPDIEHPE — — — — — — 231-245 G F SKH FLAQS F NTNED — — 260-275 QIVTVEGG L SVISPKW — — — — — — — — — — 390-407 G L SAQ Y VV L Y RNGI YSPD — 461-495 G L E Y VV F KTHHNAVSS Y IKDVF RVIPSEV L SNS Y N	186, 188, 189, 231, 232, 235, 240, 241, , 267, 268, 390, 391, 394, 395, 397, 398, 399, 403, 461, 462, 463, 464, 466, 467, 476, 477, 482, 489, 490, 493, 494	-	-

Protein name	Position and sequence	Pepsin digestion sites	Anti-digestive peptides	Allergenicity of anti-digestive peptides
P04776	— — 158-172 TPVVAVSIIDTNS L E — — — — — — — — — — — — 217-236 GGSI L SG F T L E F L EHA FSVD — 319-333 MRL RHNIGQTSSPDI — — — — — — — — — — — 347-362 L DF PAL S W L RL SAE F G — — — — — — — — 366-387 KNAM FVPHY NL NANSII Y A L NG — — — — — — — — — — 412-442 LIVPQN F VVAARSQSDN F E Y VS F KTNDTPMI — — — 468-491 SQQARQIKNNNP FK F LVPPQESQK	170, 171, 220, 221, 223, 224, 225, 226, 227, 228, 229, 232, 321, 347, 349, 352, 353, 354, 355, 357, 360, 361, 369, 374, 376, 382, 383, 384, 385, 417, 418, 428, 429, 430, 431, 433, 434, 479, 481, 482	TPVVAVSIIDTNS,	Yes
			RHNIGQTSSPDI,	No
			SQQARQIKNNNP	Yes

Protein name	Position and sequence	Pepsin digestion sites	Anti-digestive peptides	Allergenicity of anti-digestive peptides
P05046	— — — — — — — — — — — — 23-54 L V L L TSKANSAETVS F S W NK FVPKQPNMI L QG — — — — — — — — — — — — — 78-114 SS L GRA LY STPIHI WDKETGSVA S F AAS F N F T F — — — — — — 168-200 NPHIGIN-VNSIRSIKTTS W D L ANNKVAKV LIT Y — — — — 234-250 EW VRIGF SAATG L DIPG — — — — — — 253-269 HDVL S W S F ASNL PHASS	23, 24, 25, 26, 37, 38, 39, 40, 42, 51, 52, 79, 80, 83, 85, 91, 101, 102, 105, 106, 107, 108, 109, 110, 185, 186, 187, 188, 196, 199, 235, 240, 245, 246, 256, 257, 258, 259, 260, 264	NPHIGINVNSIRSIKTTS	Yes
P11827	— — — — — — — — 211-240 PFH F NSKRFQT L F KNQY GHVRV LQRF NKRS — — — — — — — — — — — — — — — — 248-291 D YRI LE F NSKPNT L LL PHHADAD Y L IVI L NGTAI L T L VNNDDRD — — — 520-535 DI F VIPAGY PVVVNAT 558-572 SKDNVISQIPSQVQE	213, 214, 221, 222, 223, 227, 232, 236, 248, 251, 253, 254, 260, 261, 263, 270, 271, 272, 275, 276, 281, 282, 283, 284, 521, 522, 528	SKDNVISQIPSQVQE	Yes

Protein name	Position and sequence	Pepsin digestion sites	Anti-digestive peptides	Allergenicity of anti-digestive peptides
P25974	— — — — — — 41-57 S F QT L F ENQNGRIRL LQ — — — — 71-85 D YRIVQ F QSKPNTI L — — — 201-226 EEQRQQEGVIVE L SKEQIRQ LSRRAK — — — — — 244-260 SRNPIY SNN F GK F FEIT — — — — — — — — 271-291 DI F L SSVDINEGA L LL PHF NS — — — — 335-349 SEDDV F VIPAAY P FV	41, 42, 44, 45, 46, 55, 71, 76, 77, 84, 212, 213, 220, 249, 252, 253, 255, 256, 272, 273, 274, 283, 284, 286, 289, 339, 340, 346, 347	EEQRQQEGVIVE	No
P26987	— — — 19-34 L Y KA LVTDADNVIPKA — — — — — 49-74 GPGTIKKITF L EDGETK F VL HKIESI — — — 106-120 AGPNGGSAGK L TVK Y — — — — — — — — — 138-155 AKA DA L F KAIEA Y L L AHP	19, 20, 22, 58, 59, 65, 66, 68, 115, 116, 119, 142, 143, 144, 149, 150, 151, 152	LVTDADNVIPKA	Yes

Fig. 1

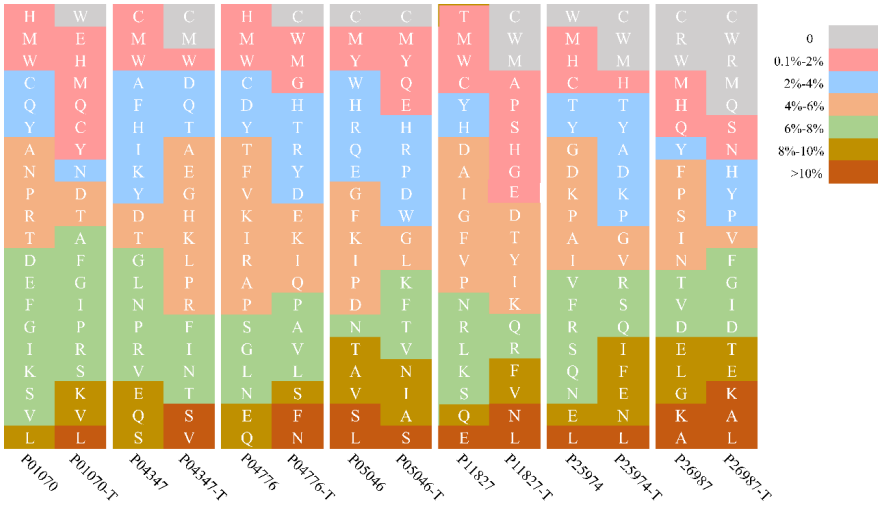
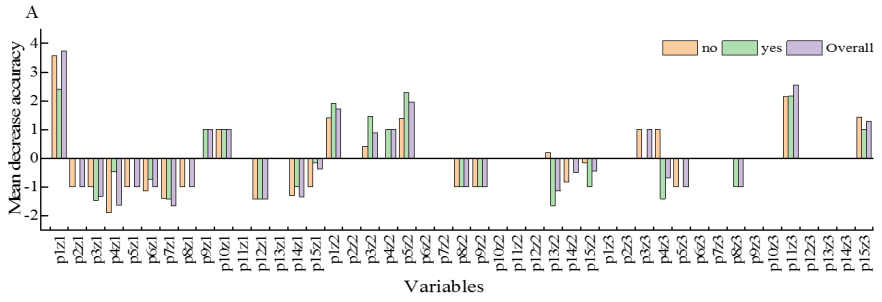
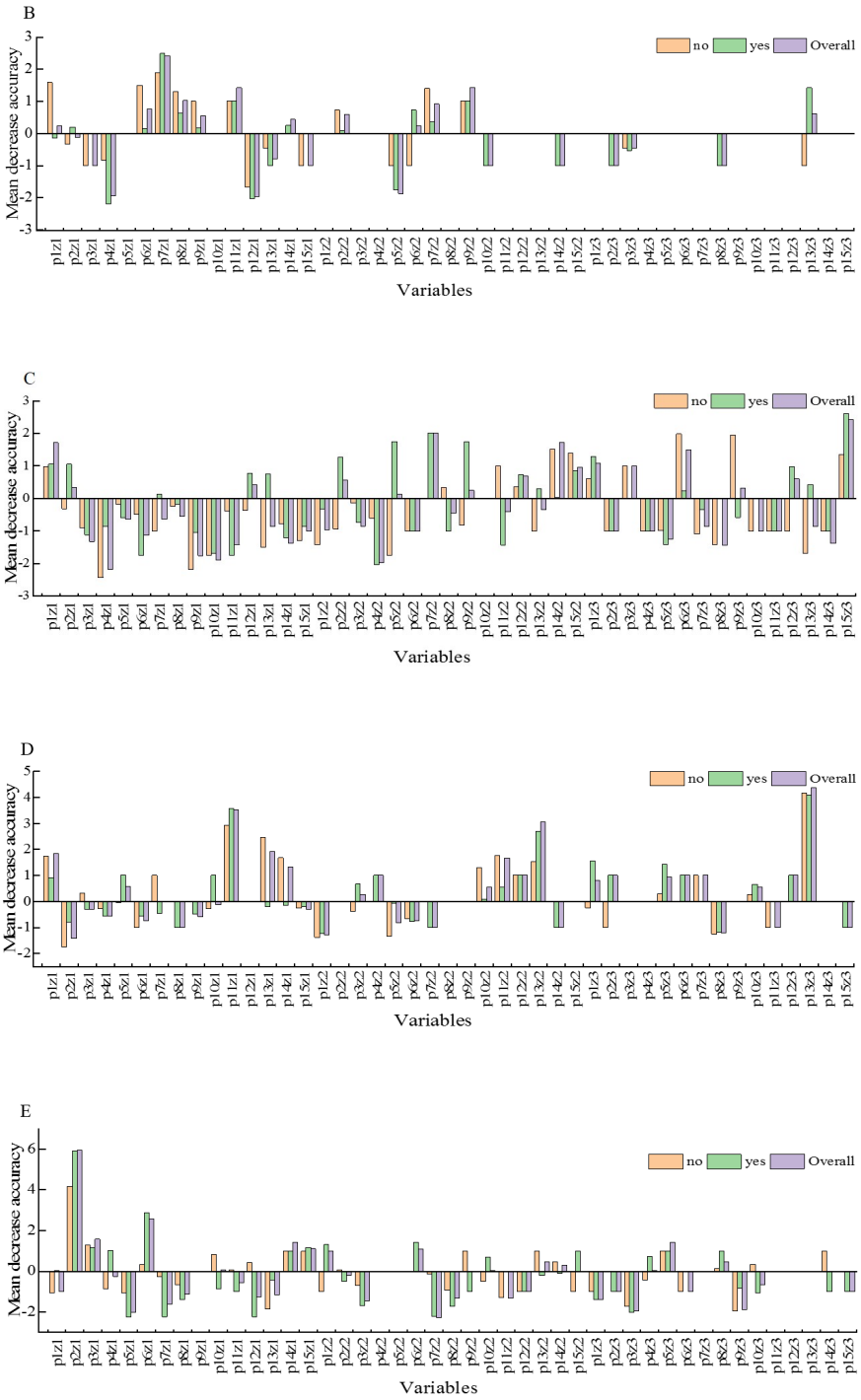


Fig. 1 Analysis of the amino acid composition of T cell epitopes of seven soybean allergens.

(P01070: Amino acid composition analysis of P01070 protein; P01070-T: Amino acid composition analysis of T cell epitopes of P01070 protein. P04347: Amino acid composition analysis of P04347 protein; P04347-T: Amino acid composition analysis of T cell epitopes of P04347 protein. P04776: Amino acid composition analysis of P04776 protein; P04776-T: Amino acid composition analysis of T cell epitopes of P04776 protein. P05046: Amino acid composition analysis of P05046 protein; P05046-T: Amino acid composition analysis of T cell epitopes of P05046 protein. P11827: Amino acid composition analysis of P11827 protein; P11827-T: Amino acid composition analysis of T cell epitopes of P11827 protein. P25974: Amino acid composition analysis of P25974 protein; P25974-T: Amino acid composition analysis of T cell epitopes of P P25974 protein. P26987: Amino acid composition analysis of P26987 protein; P26987-T: Amino acid composition analysis of T cell epitopes of P26987 protein.)

Fig. 2





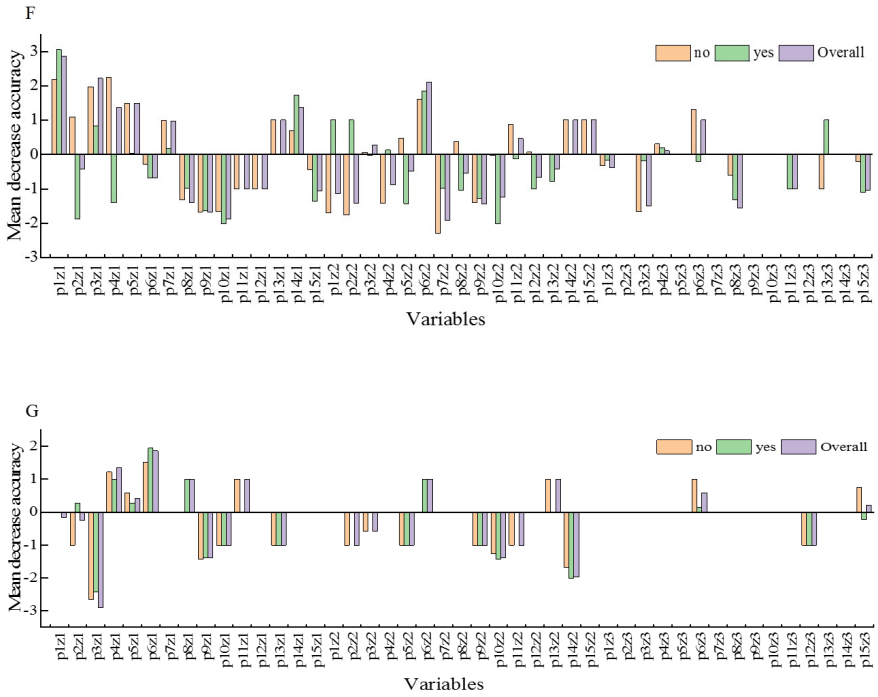


Fig. 2 Mean decrease accuracy of the 3-z scale models of the (A) P01070 T cell epitopes, (B) P04347 T cell epitopes, (C) P04776 T cell epitopes, (D) P05046 T cell epitopes, (E) P11827 T cell epitopes, (F) P25974 T cell epitopes, and (G) P26987 T cell epitopes.