

Sequence analysis of digestion-resistant peptides may be an efficient strategy for studying the linear epitopes of Jug r 1, the major walnut allergen

Xiaoya Guo^a, Songsong Jiang^{a,c}, Xinrui Li^{a,d}, Shuai Yang^a, Lei Cheng^a, Jinyu Qiu^b, Huilian Che^{a,*}

^a Beijing Advanced Innovation Center for Food Nutrition and Human Health, College of Food Science and Nutritional Engineering, China Agricultural University, No. 17 Qinghua East Road, Haidian District, Beijing, PR China

^b College of Pharmacy, Jinan University, No. 601 Huangpu Road, Tianhe District, Guangzhou, PR China

^c College of Food Science and Engineering, Yangzhou University, No. 169 Huayang West Road, Hanjiang District, Yangzhou, PR China

^d Biocells (Beijing) Biotech Co., Ltd., Haiying Road, Fengtai District, Beijing, PR China

ARTICLE INFO

Keywords:

Linear epitopes

Jug r 1

Bioinformatics analysis

Overlapping peptide synthesis

Digestion-resistant peptides

ABSTRACT

Jug r 1, the major allergen of walnut, triggers severe allergic reactions through epitopes. Hence, research on the efficient strategy for analyzing the linear epitopes of Jug r 1 are necessary. In this work, bioinformatics analysis was used to predict the linear epitopes of Jug r 1. Overlapping peptide synthesis was used to map linear epitopes. *In vitro* simulated gastrointestinal digestion and HPLC-MS/MS were used to identify digestion-resistant peptides. The results showed that six predicted linear epitopes were AA28-35, AA42-49, AA55-62, AA65-73, AA97-104, and AA109-121. AA16-30 and AA125-139 were identified by the sera of walnut allergic patients. Five digestion-resistant peptides were AA19-33, AA40-45, AA54-74, AA96-106, and AA117-137. The predicted results only included one of the linear epitopes identified by sera, while the digestion-resistant peptides covered all. Therefore, the digestion-resistant property of food allergens may be a promising direction for studying the linear epitopes of Jug r 1.

1. Introduction

Food allergy, one of the allergic diseases, can be life-threatening (Kimber & Dearman, 2001). In recent years, the incidence of food allergy has been increasing (Jennifer et al., 2010). Walnut allergy is a common type of tree nut allergy responsible for severe anaphylactic reactions (Aalberse, 1997). So far, six walnut allergens have been identified, including Jug r 1, Jug r 2, Jug r 3, Jug r 4, Jug n 1, and Jug n 2 (Roux, Teuber, & Sathe, 2003). Among these allergens, Jug r 1 is recognized by more than 80% of the walnut allergic patient sera (Teuber, Dandekar, Peterson, & Sellers, 1998), suggesting that it is the major allergen of walnut. Jug r 1 can induce food allergy due to its structure, which can be recognized as epitopes. Epitopes bind with antigen-specific IgE leading to anaphylaxis (Liu & Sathe, 2018). Downs et al. (2014) found that both the non-reduced and reduced 2S albumin from English walnut could bind with serum IgE of walnut allergic subjects. This research proved that epitopes can be either linear sequences of amino acids (linear epitopes) or three-dimensional motifs (conformational epitopes). In this study, we focus on the linear epitopes

of Jug r 1.

There are only a few reports on the linear epitopes of Jug r 1, although its amino acid sequence is well known. Until now, linear epitopes can be found using different methods. Bioinformatics software is an emerging approach to predict linear epitopes. This approach is based on antigenic properties of amino acid sequences (Mari et al., 2006, He & Tao, 2015). However, the predicted linear epitopes are not equal to real linear epitopes. To validate the accuracy of predicted results, overlapping peptide synthesis is most commonly used (Liu & Sathe, 2018). This approach implies a synthesis of short peptides covering the entire protein sequence with a shift of 3–5 amino acids as described by Jung and Beck-Sickinger (2010). The sera of allergic patients were then used to map linear epitopes from these peptides. Pahr et al. (2014) used this technique to synthesize overlapping peptide fragments of wheat food allergen Tri a 37 and screen its linear epitopes. However, the synthesis of large numbers of peptides is relatively time-consuming and labour-intensive. In addition, a large volume of sera is required. Recently, peptide microarray, a novel assay that facilitates high-throughput screening of peptides with a small quantity of sample, has been used to

* Corresponding author.

E-mail address: chehuilian@cau.edu.cn (H. Che).

<https://doi.org/10.1016/j.foodchem.2020.126711>

Received 13 August 2019; Received in revised form 27 March 2020; Accepted 29 March 2020

Available online 30 March 2020

0308-8146/© 2020 Elsevier Ltd. All rights reserved.

map linear epitopes as reported by Rosenfeld et al. (2012). They assessed linear epitopes of peanut using this assay. Although this assay enables simultaneous screening of thousands of target peptides by using small volumes of sera, it is relatively expensive.

Food allergy is an immune response that occurs in intestinal tract (Macdonald & Monteleone, 2005). Bredehorst and David (2001) reported that digestibility the key factors affecting the allergenic potential of allergens. Therefore, the World Health Organization (2001) considers protein digestibility as one of the main criteria for assessing the allergenic potential of genetically modified organisms. Jug r 1, as a 2S albumin containing multiple disulfide bonds, is relatively stable during heating process and digestion (Breiteneder & Radauer, 2004). Hence, most of the linear epitopes of Jug r 1 may be digestion-resistant. There are few reports associate the digestion-resistant property of food allergens with linear epitopes. Moreno, Mellon, Wickham, Bottrill, and Mills (2005) identified digestion-resistant peptides of the major 2S albumin allergen from Brazil nuts Ber e 1. They found that these peptides may contain conformational and linear epitopes. Therefore, we hypothesized that the digestion-resistant property of food allergens is a possible direction for studying the linear epitopes of Jug r 1.

In the present study, we compared the differences in efficiency between conventional methods (bioinformatics analysis, overlapping peptide synthesis) and methods based on the digestion-resistant property of food allergens (*in vitro* simulated gastrointestinal digestion, HPLC-MS/MS) to provide an efficient strategy for analyzing the linear epitopes of Jug r 1.

2. Material and methods

2.1. Material

Jug r 1 was prepared earlier by our group to a purity of 90%, as assessed by LC-MS/MS. Jug r 1 was separated and purified from walnut (*J. regia* L., produced in Tianjin, China) through ammonium sulfate fractionation precipitation, Superdex 200 gel filtration chromatography, and Q-Sepharose Fast Flow anion exchange chromatography.

2.2. Bioinformatics analysis

2.2.1. Amino acid sequence analysis

The complete amino acid sequence of Jug r 1 was obtained from the NCBI database (Lin et al., 2014). The amino acid sequence was analyzed using SignalP 4.1 Server and ProtParam tool in ExPASy (Sigrist et al., 2013).

2.2.2. Secondary structure analysis

In the DNASTar Protean system, the secondary structure analysis of Jug r 1 was performed using the methods of Chou-Fasman and Garnier-Robson (Shanti, Martin, Nagpal, Metcalfe, & Rao, 1993; Leung et al., 1994). The proportions of α -helix, β -corner, β -folding, and random coil were analyzed.

2.2.3. Tertiary structure analysis

In the Protein Data Bank, BLAST was used to search a known protein as a template (Lin et al., 2014). The template should possess a high degree of sequence homology with Jug r 1. Therefore, Ber e 1 was selected for its approximately 45% identity with Jug r 1. Thus, we built a rather accurate three-dimensional (3D) model using the alignment mode in the SWISS-MODEL website of the 2S albumin as a template (Arnold, Bordoli, Kopp, & Schwede, 2006). Furthermore, the 3D model of Jug r 1 was obtained after molecular dynamics simulation and energy minimization. The visualization software PyMol was used to view the position of predicted linear epitopes and digestion-resistant peptides of Jug r 1 in the tertiary structure.

2.2.4. Linear epitope prediction

Immunoinformatic software DNASTar Protean system and online tools, including Bioinformatics Predicted Antigenic Peptides (BPAP) system and BepiPred 1.0 Server were used to predict the linear epitopes of Jug r 1 (Yang & Yu, 2009).

In the DNASTar Protean system, four properties of the amino acid sequence were considered as the parameters for epitope prediction, including hydrophilicity, flexibility, accessibility, and antigenicity. The hydrophilicity prediction was performed according to the methods described by Hopp and Woods (1981), and Kyte and Doolittle (1982). Moreover, the properties of flexibility, surface accessibility, and antigenicity were analyzed using the methods described by Karplus and Schulz (1985), Emini, Hughes, Perlow, and Boger (1985), and Jameson and Wolf (1988), respectively. Based on the results of the prediction methods mentioned above, the peptide regions with appropriate hydrophilicity, high flexibility, good surface accessibility, and high antigenic index were predicted as linear epitopes by DNASTar. The BPAP system and the BepiPred 1.0 server were based on the physicochemical properties of amino acids, such as hydrophilicity, flexibility, accessibility, turns, and exposed surface, to predict linear epitopes.

2.3. Overlapping peptide synthesis

2.3.1. Peptide synthesis

Peptides of 15 amino acids in length with an offset of 3, covering the complete sequence of Jug r 1, were commercially synthesized as linear epitopes are typically 8–15 amino acids long (Lin et al., 2009). The purity of these peptides was higher than 90%, as assessed by HPLC. The molecular mass of the synthetic peptides was confirmed by electrospray ionization mass spectrometry (ESI-MS). The peptides were stored at -20°C until use.

2.3.2. Human sera

Sera from six Chinese walnut allergic patients with a convincing history of walnut anaphylaxis were used to map the linear epitopes of Jug r 1. The characteristic information of walnut allergic patients is shown in Table S1. Blood samples were centrifuged at 3000g and 4°C for 10 min following coagulation for 1 h at room temperature to obtain sera. The sera were stored at -80°C before analysis.

2.3.3. Ethical approval

The study protocol was approved by the Human Research Ethics Committee of China Agricultural University (Approval Number: 2015017).

2.3.4. Western blotting

Western blotting was performed to detect the binding of Jug r 1 with specific IgE in allergic patient sera (Liu et al., 2017). The sera of four non-allergic patients were mixed as a negative control. Proteins were subjected to 15% sodium dodecylsulphate–polyacrylamide gel electrophoresis (SDS-PAGE), and the separated proteins were transferred to polyvinylidene fluoride (PVDF) membranes (MilliporeSigma, Merck KGaA, Darmstadt, Germany). The membranes were blocked in Tris-buffered saline containing 0.1% Tween 20 (Xilong Science Co., Ltd, Shantou, Guangdong Province, China) and 5% non-fat milk powder (Inner Mongolia Yili Industrial Group Co., Ltd, Inner Mongolia Autonomous Region, China), and then incubated with sera overnight at 4°C . Non-fat milk powder was used to block non-specific binding sites. Next, the membranes were incubated with specific peroxidase-conjugated goat anti-human IgE antibody (Sigma-Aldrich Inc., St. Louis, Missouri, USA) for 2 h at room temperature. Finally, the membranes were incubated in enhanced chemiluminescence (ECL) reagent (Pierce ECL Western blotting Substrate, Thermo Fisher Scientific Inc., Waltham, Massachusetts, USA) and exposed to film.

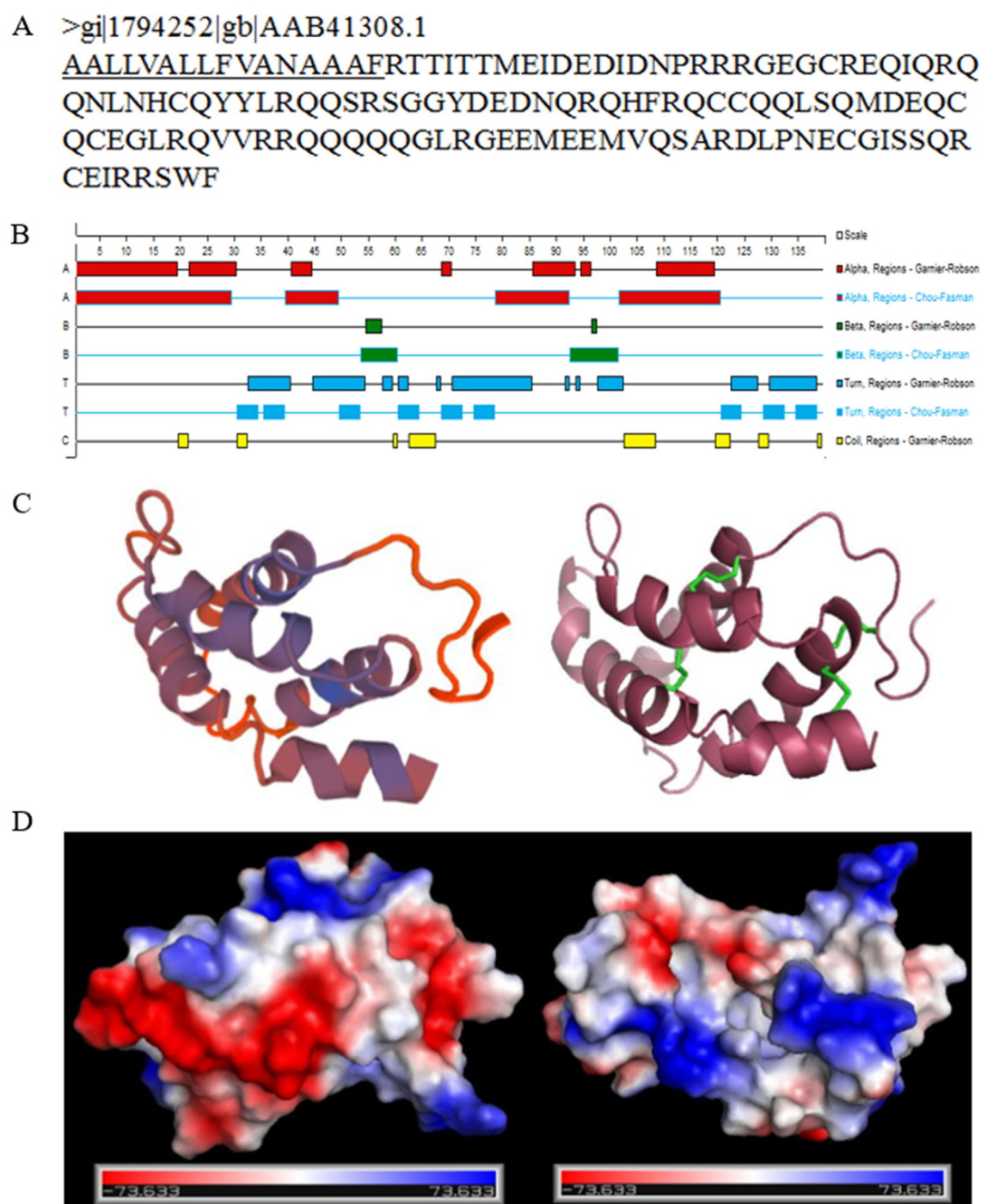


Fig. 1. Amino acid sequence, secondary structure, and tertiary structure analysis of Jug r 1. **A:** Amino acid sequence of Jug r 1 (underline is the predicted signal peptide of Jug r 1). **B:** Secondary structure prediction of Jug r 1. **C:** Ribbon model of Jug r 1. The elevation view is on the left. The rear-side elevation is on the right, and the green labels represent disulfide bonds. **D:** Surface potential distribution of Jug r 1. The elevation view is on the left. The rear-side elevation is on the right. The blue part represents the positive charge area on the molecular surface, and the red part represents the negative charge area on the molecular surface. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.3.5. Dot-blot

Dot blot experiments were performed to map the peptides with serum IgE (Bertelloni et al., 2018). Nitrocellulose membranes (Whatman, Maidstone, Kent, England) were immersed in double-distilled water for 10 min. When the membranes were damp-dry, the peptides dissolved in carbonate buffered saline (CBS, pH 9.6) were spotted on the membranes and left to dry at room temperature (4 mg for each dot). The spotted membranes were incubated for 2 h in 0.01 M pH 7.4 phosphate-buffered saline containing 0.1% Tween 20 (PBST) and

5% non-fat milk powder. The membranes were then incubated with human sera (diluted 25 times with PBST containing 2.5% non-fat milk powder). After overnight incubation at room temperature, the membranes were washed for 5 min with PBST and further incubated in PBST containing 2.5% non-fat milk powder and peroxidase-conjugated goat anti-human IgE antibody at 37 °C for 1 h. The titer of the antibody was 1:8000, which was obtained according to the supplier's suggested working dilution. Afterwards, the membranes were washed three times with PBST, incubated in ECL reagent for 1 min, removed excess reagent

solution from the surface, and then carefully exposed to X-ray film.

2.4. *In vitro simulated gastrointestinal digestion*

In vitro simulated gastrointestinal digestion was performed using the method of Toomer, Do, Pereira, and Williams (2013) to test the digestion stability of Jug r 1 and obtain digestive products. *In vitro* simulated gastric fluid digestion and *in vitro* simulated intestinal fluid digestion were performed respectively.

2.4.1. *In vitro simulated gastric fluid digestion*

Simulated gastric fluid (SGF) was prepared as described by Toomer et al. (2013) with some modifications: 100 mL double-distilled water comprised 131.6 mg pepsin (Sigma-Aldrich Inc., St. Louis, Missouri, the USA) and 0.2 g NaCl (adjusting pH to 1.2 with HCl, pepsin activity ≥ 2000 U/mg).

Six 1.5 mL centrifuge tubes containing 190 μ L SGF were incubated in a 37 °C water bath for 5 min. Five tubes were then added with Jug r 1 (50 μ g) and incubated in a 37 °C water bath. At 0 s, 15 s, 2 min, 30 min, and 60 min, one of these tubes was removed, added to 70 μ L sodium bicarbonate solution (0.2 mol/L) and placed in an ice bath to terminate the reaction. The remaining tube was the pepsin control group. It was added with the solvent of Jug r 1. The reaction was terminated by adding 70 μ L sodium bicarbonate solution (0.2 mol/L), and placing the tube in an ice bath. The SGF digestive products and pepsin control group were subsequently stored at -20 °C until SDS-PAGE was performed.

Pepsin-free SGF was used to estimate the thermal stability of Jug r 1 under digestive conditions. We repeat the same steps mentioned above.

2.4.2. *In vitro simulated intestinal fluid digestion*

Simulated intestinal fluid (SIF) was prepared as described by Toomer et al. (2013) with some modifications: 100 mL double-distilled water comprised 1 g pancreatin (Sigma-Aldrich Inc., St. Louis, Missouri, USA) and 0.7 g KH₂PO₄ (adjusting pH to 7.5 with NaOH, pancreatin needs to meet the requirement of digesting casein which is 25 times its mass within 60 min at 40 °C (pH7.5)). Repeat the same steps mentioned above, except that the reaction was terminated by boiling in a water bath for 5 min.

2.5. *Detection of the digestion-resistant peptide sequences of Jug r 1*

The digestive products of Jug r 1 were separated using nano-upgrading HPLC system Easy-nLC 1000. Mobile phase: 0.1% formic acid-water solution (A) and 0.1% formic acid-acetonitrile solution (B). The chromatographic column was balanced with 95% liquid A. Samples were loaded through an automatic sampler to a pre-column C18 trap column (3 mm, 0.10 $\times$ 20 mm), then separated using an analytical column C18 column (1.9 mm, 0.15 $\times$ 120 mm), with a flow rate of 600 nL/min. Each sample was separated by capillary HPLC and analyzed by Q-Exactive mass spectrometer.

PFind 2.6 software was used to identify the sequence of Jug r 1 in the digestive products via tandem mass spectrometry, which we called database searching. The database searching parameters were as follows: enzyme: pepsin or pancreatin; max missed cleavages: 2; peptide mass tolerance: ± 15 ppm; fragment mass tolerance: 0.02 Da; peptide confidence: high; peptide length: > 4 . The filtration parameters were as follows: peptide FDR ≤ 0.01 .

3. Results and discussion

3.1. *Characteristics of amino acid sequence, secondary structure, and 3D model of Jug r 1*

For predicting the linear epitopes of Jug r 1, it is necessary to know its amino acid sequence. The amino acid sequence of Jug r 1 was

analyzed using SignalP 4.1 Server. As shown in Fig. 1A, the first 16 amino acid residues in the N-terminal are signal peptides, so the post-translational mature protein sequence does not include this region. The sequence of the non-signal peptides was analyzed using the ProtParam tool in ExPASy. The theoretical isoelectric point of Jug r 1 is 5.7, and the molecular weight is 14816.3 Da.

The secondary structure plays an important role in epitope formation. It is generally known that α -helices and β -sheets are located inside with high chemical bond energy. Therefore, it is difficult to form linear epitopes in these regions. The β -turns and random coils are mostly located on the surface with a loose structure, which makes it easy to bind with antibodies and form linear epitopes. The DNASTar Protean system was used to analyze the secondary structure of Jug r 1. As shown in Fig. 1B, β -turn is dominant in the secondary structure, suggesting that these regions are likely to form linear epitopes. The appearance of α -helix is also frequent, but these regions cannot form linear epitopes as they are difficult to bind with antibodies.

The 3D model of Jug r 1 was constructed using the Swiss-model server. As shown in Fig. 1C (left), the α -helix is dominant in the tertiary structure, which is consistent with the predicted results of DNASTar. There are disulfide bonds connecting the five main α -helices in the molecular structure (Fig. 1C right). The disulfide bonds and α -helices play important roles in maintaining the stability of the space structure. Therefore, Jug r 1 has a steady spherical structure. In addition, as shown in Fig. 1D, Jug r 1 has an extensive charge distribution, strong polar, and good solvent accessibility. These characteristics provide favorable conditions for epitope formation.

3.2. *Six linear epitopes of Jug r 1 were predicted by bioinformatics analysis*

Surface accessibility and fragment flexibility are important features for predicting linear epitopes (Karplus and Schulz, 1985; Emini et al., 1985). Moreover, the regions with high hydrophilicity also provide strong evidence for linear epitope identification (Hopp & Woods, 1981; Kyte & Doolittle, 1982). The DNASTar Protean system is an immunoinformatic tool that use these three variables and antigenicity to predict linear epitopes. As shown in Fig. 2A, the hydrophilic regions occupying most of the sequence indicate the high hydrophilicity of Jug r 1. In addition, the frequent appearance of surface accessibility and flexibility regions indicates that the Jug r 1 sequence was flexible, stretched, and easily exposed outside. The high antigenic index in the complete sequence suggests the high tendency of the Jug r 1 protein to form linear epitopes (Fig. 2A). Based on these sequence properties, the final predicted linear epitopes of DNASTar are AA28-35, AA42-49, AA55-62, AA65-73, AA97-104, and 109-121 as shown in Table S2.

The BPAP system and the BepiPred 1.0 server were also used to predict the linear epitopes of Jug r 1. The predicted results of the three aforementioned tools are shown in Table S2. There are many overlapping parts in the linear epitopes predicted by the three tools. The linear epitopes predicted by at least two tools were considered as the final results, and the controversial results were discarded. The final predicted linear epitopes were 6 peptides, including AA28-35, AA42-49, AA55-62, AA65-73, AA97-104, and AA109-121.

PyMOL was used to locate the predicted linear epitopes of Jug r 1 in the tertiary structure. As the N-terminal region of the tertiary structure differs significantly from the modeling template Ber e 1, it is purposefully omitted in the model. Accordingly, the model of Jug r 1 starts at residue 35, and the first predicted linear epitope (AA28-35) is eliminated. As shown in Fig. 2B, three predicted linear epitopes are at the joint of α -helix and random coil with a high possibility of binding with serum IgE. The remaining two predicted linear epitopes are at α -helix, which are difficult to bind with serum IgE. All five predicted epitopes were exposed with high surface accessibility (Fig. 2C), which might promote the interaction with serum IgE. Therefore, the five predicted epitopes might be the potential linear epitopes of Jug r 1.

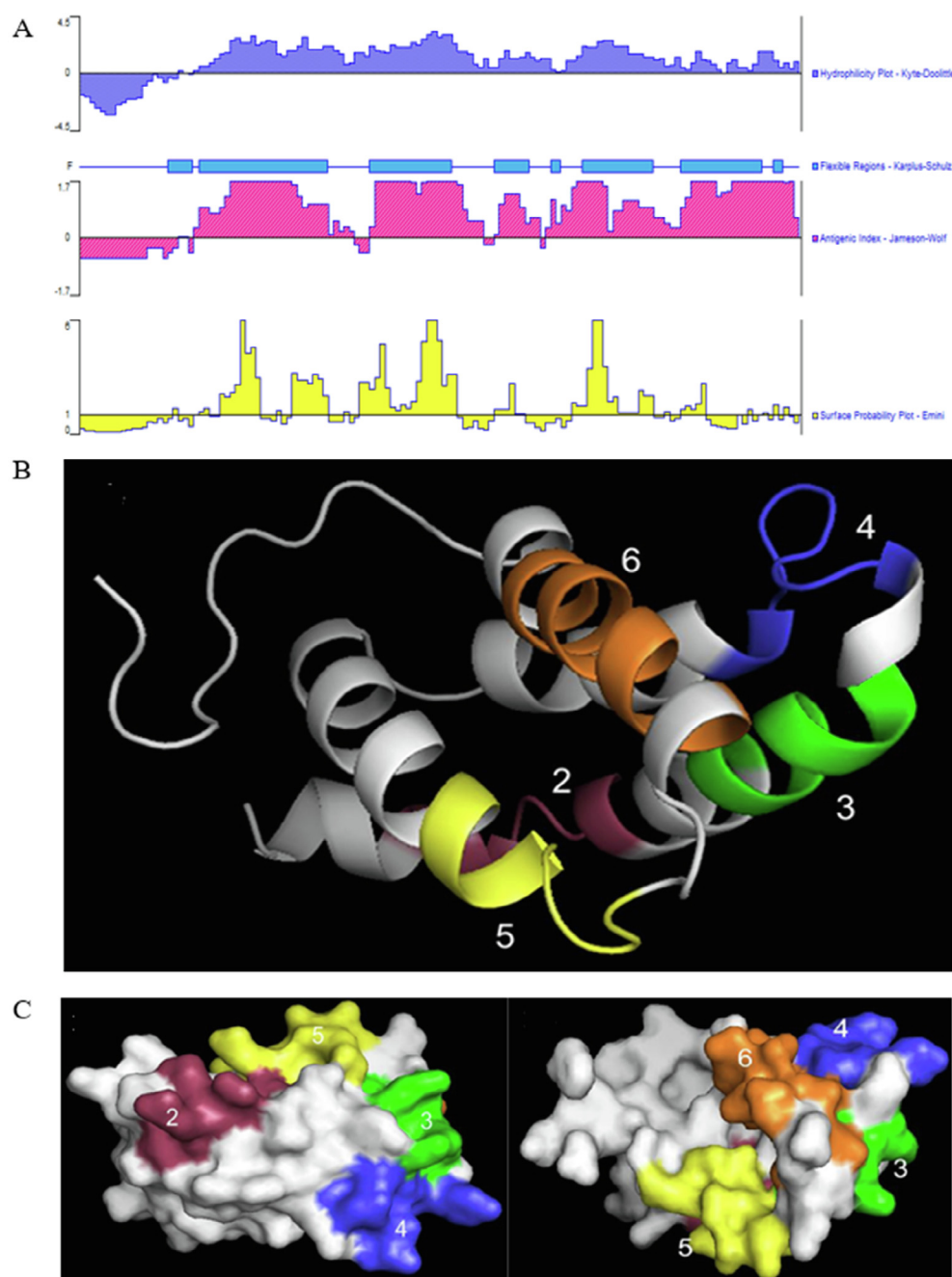


Fig. 2. Prediction of the linear epitopes of Jug r 1. A: Amino acid sequence properties of Jug r 1 analyzed by DNASTar. B: Location of the predicted linear epitopes in the ribbon model of Jug r 1. C: Location of the predicted linear epitopes in the surface potential distribution model of Jug r 1.

3.3. Two linear epitopes of Jug r 1 were screened by overlapping peptide synthesis

Western blotting was conducted with the sera of six Chinese walnut allergic patients to validate the IgE binding with Jug r 1. As shown in Fig. 3A, all six sera from allergic patients demonstrated strong binding with Jug r 1, and the serum from the fifth patient revealed the most potent binding capacity, suggesting the existence of Jug r 1-specific IgE in all six sera. It also illustrates that walnut allergy in these six patients is induced by Jug r 1, confirming that Jug r 1 is the major walnut allergen in China. As Jug r 1 was prepared by our group with a purity of 90%, the binding bands at about 25 kDa and 10 kDa shown in Patients 4, 5, and 6 may be the bands binding with other walnut allergens.

To validate the linear epitopes predicted by bioinformatics analysis,

dot-blot immunoassay was conducted with the 44 synthetic peptides to map linear epitopes. This assay was based on the reactivity between Jug r 1 and serum IgE in walnut allergic subjects. As shown in Fig. 3B, among the 44 synthetic peptides, four reacted to all six sera from walnut allergic patients. These are peptide 1 (AA1-15), peptide 2 (AA4-18), peptide 6 (AA16-30), and peptide 44 (AA125-139). As mentioned in 3.1, the 16 amino acid residues in front of the N-terminal are signal peptides and the post-translational mature protein sequence does not include this region. Therefore, peptides 1 and 2 could not be linear epitopes of Jug r 1. In the present study, we consider AA16-30 and AA125-139 as real linear epitopes of Jug r 1, and they were used to evaluate the accuracy of the predicted results. We found that these linear epitopes identified by sera were not in line with the predicted results by bioinformatics. The differences between them require further

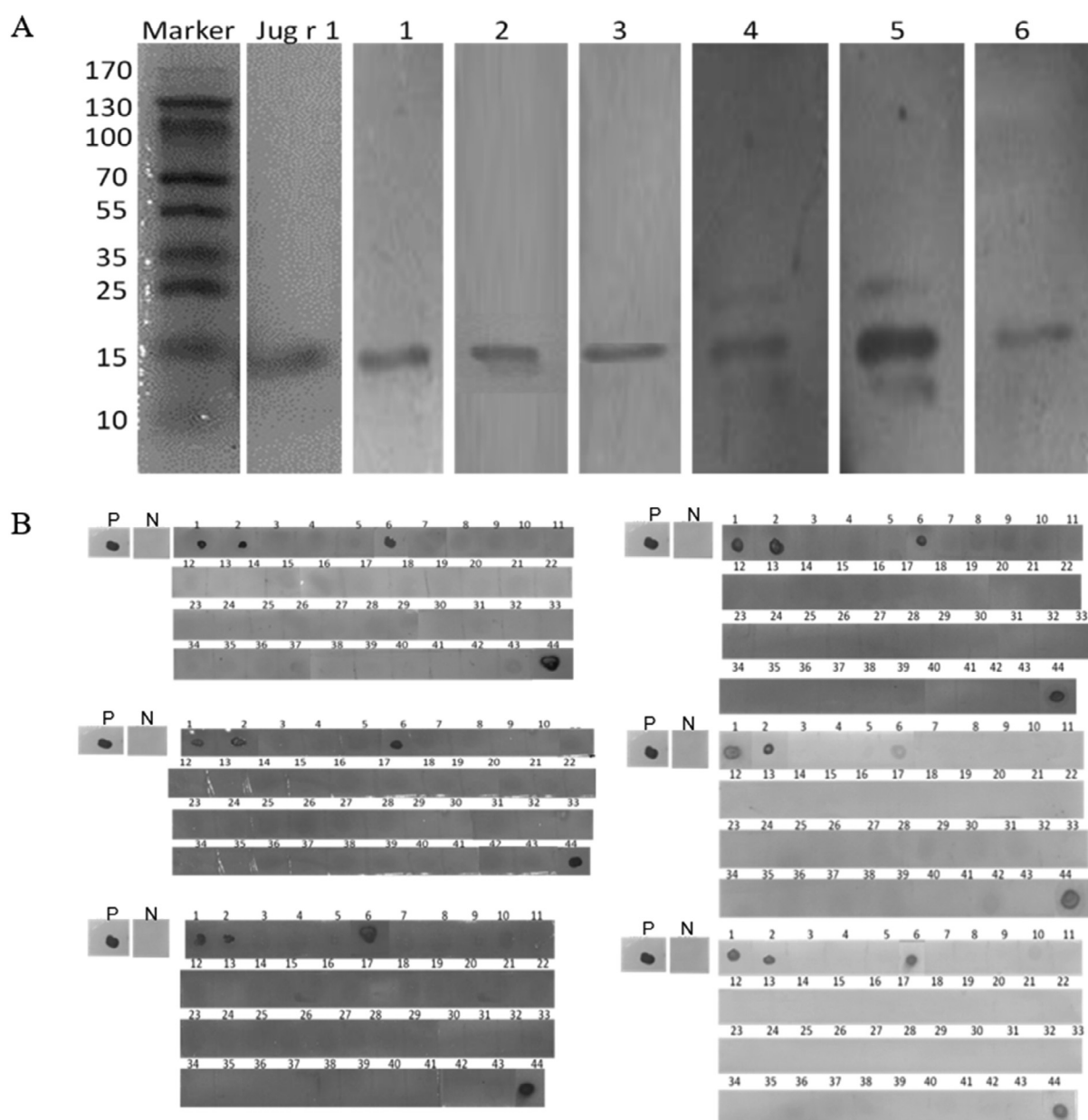


Fig. 3. Screen of real linear epitopes of Jug r 1. A: Western blot results of sera from six Chinese walnut allergic patients with Jug r 1. The lane labeled “Jug r 1” is an SDS-PAGE gel image. B: Dot-blot results of sera from six Chinese walnut allergic patients with Jug r 1 (P: positive, N: negative).

study.

3.4. Jug r 1 is stable during digestion

Proteins that cannot be detected after two min of digestion are unstable, proteins that cannot be detected between two and 15 min of digestion are partially stable, while proteins that can still be detected after 60 min of digestion are stable (Goodman & Hefle, 2005). Therefore, the maximum digestion time was set to 60 min in the present study to evaluate the digestive stability of Jug r 1.

The result of simulated gastric digestion of Jug r 1 is shown in Fig. 4A (left). Jug r 1 maintains integrity only at 0 s of digestion. It was decomposed into fragments gradually as digestion proceeded. Two low-molecular-weight bands appeared at 15 s of digestion. At 30 min of digestion, Jug r 1 was completely decomposed, but its subunit bands remained intact at the end of digestion. This means that the subunit is stable to pepsin digestion. The structure of Jug r 1 can explain this result. It belongs to the 2S albumin family, which has a compact

structure. Disulfide bonds of 2S albumin are broken under acid or enzymatic hydrolysis conditions (Moreno et al., 2005). Then, it is decomposed into a large subunit and a small subunit. The structure of subunits is very stable, thus they are not further digested or decomposed. It is an important immunological characteristic of Jug r 1. The result of simulated intestinal digestion of Jug r 1 is shown in Fig. 4A (right). It is also gradually decomposed into fragments as digestion proceeded. At 30 min of digestion, Jug r 1 was completely digested and could not be detected by SDS-PAGE. This means that Jug r 1 is partially stable to pancreatin digestion. The results indicate that Jug r 1 is stable during digestion.

3.5. Five digestion-resistant peptides of Jug r 1 were identified

HPLC-MS/MS was used to identify the digestion-resistant peptides of Jug r 1 from the digestive products. After digestion with pepsin for 60 min, sixty-eight secondary mass spectra were detected. Seven fragments belonging to Jug r 1 were identified by pFind, covering 38% of

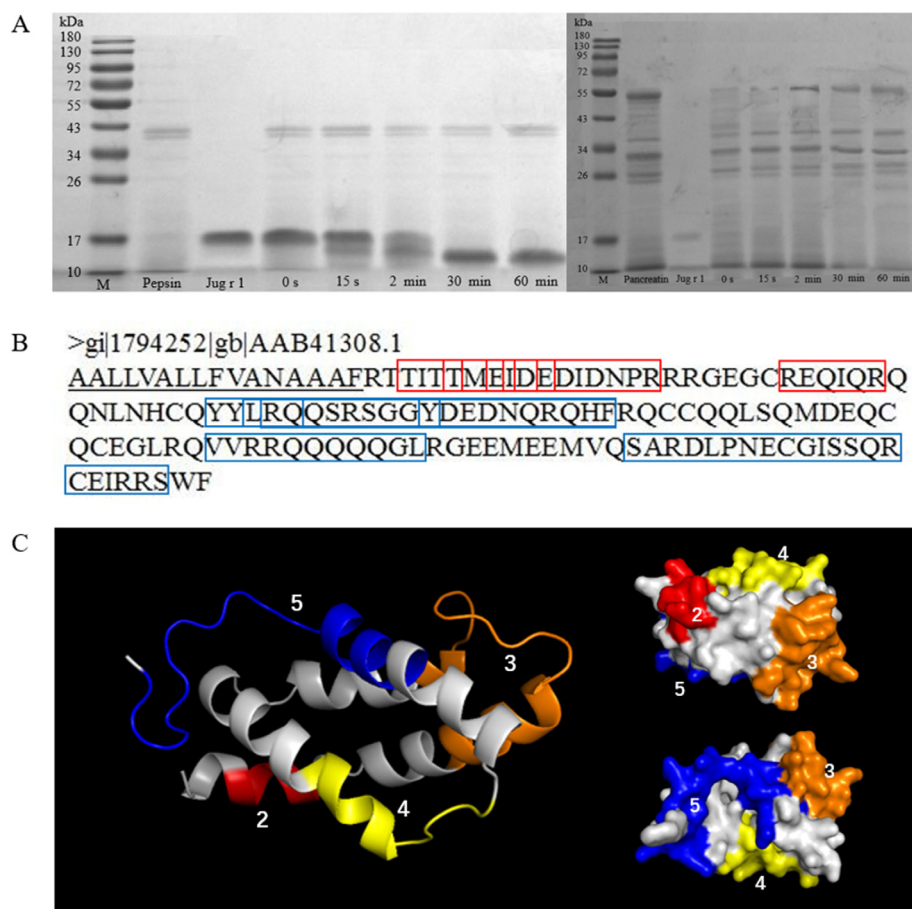


Fig. 4. Identification of the digestion-resistant peptides of Jug r 1. A: SDS-PAGE electrophoresis of Jug r 1 digested by pepsin (left) and pancreatin (right) (M: marker). B: Location of digestion-resistant peptides in the amino acid sequence of Jug r 1 (the blue box: fragments resistant to pepsin digestion; the red box: fragments resistant to pancreatin digestion). C: Location of digestion-resistant peptides in the tertiary structure of Jug r 1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the amino acid sequence (Table S3). After digestion with pancreatin for 60 min, eighty-seven secondary mass spectra were detected. Nine fragments belonging to Jug r 1 were identified by pFind, covering 15% of the amino acid sequence (Table S4). The peptide 1, peptide 2, peptide 8 and peptide 9 are all modified peptides, so the observed molecular weight is slightly higher than the theoretical molecular weight.

These digestion-resistant peptides were marked in the amino acid sequence of Jug r 1. As shown in Fig. 4B, the fragments resistant to pepsin digestion are in the blue box, and the fragments resistant to pancreatin digestion are in the red box. Overall, these digestion-resistant peptides were distributed at the N-terminal, middle, and C-terminal of the Jug r 1 amino acid sequence, covering about 53% of the sequence. Five digestion-resistant overlapping peptides appeared at AA54-74 after digestion with pepsin, and eight digestion-resistant overlapping peptides appeared at AA19-33 after digestion with pancreatin. This proved the digestion-resistant properties of these regions. Whether linear epitopes exist in these regions still requires further analysis.

PyMOL was used to locate the digestion-resistant peptides of Jug r 1 in the tertiary structure. The digestion-resistant peptides were grouped into five non-overlapping peptides: AA19-33, AA40-45, AA54-74, AA96-106, and AA117-137. As the model of Jug r 1 started at residue 35, except AA19-33, the remaining four digestion-resistant peptides were marked in the tertiary structure. As shown in Fig. 4C, three digestion-resistant peptides are at the joints of α -helices and random coils with a high possibility of binding with serum IgE. One remaining peptide is at α -helix, which is difficult to bind with serum IgE. In addition, they were almost exposed to high surface accessibility (Fig. 4C), which might promote the interaction with serum IgE. Therefore, they are all possible to be linear epitopes of Jug r 1. According to the results of overlapping peptide synthesis, the digestion-resistant peptides cover

all the linear epitopes of Jug r 1 recognized by the sera of Chinese walnut allergic patients (AA16-30 and AA125-139). This suggests that sequence analysis of digestion-resistant peptides may be a promising strategy for studying the linear epitopes of Jug r 1 compared with the conventional methods.

3.6. Discussion

Antigen-specific IgE plays a critical role in food allergy, which is usually accompanied by serious clinical symptoms (Sun, Zhou, Sun & Che, 2015). Thus, identification of IgE-binding epitopes (including linear epitopes and conformational epitopes) is important for better understanding the mechanisms of food allergy. In the present study, we focused on the identification of linear epitopes. First, we used three immunoinformatic tools to predict the linear epitopes of Jug r 1. Next, overlapping peptide synthesis was used to map the linear epitopes among the 44 synthetic peptides. Six predicted linear epitopes of Jug r 1 were AA28-35, 42 AA-49, AA 55-62, AA 65-73, AA 97-104, and AA 109-121. Although immunoinformatic tools are commonly used and have been reported to be relatively successful, these tools based on amino acid propensity profiles cannot be used reliably to predict linear epitopes (Blythe & Flower, 2010). When the 3D structure of the antigen is available or can be reliably predicted, it can be used to increase the accuracy of the predicted linear epitopes. Therefore, we constructed a 3D structure model of Jug r 1, and located the predicted epitopes within it. The results showed that the predicted linear epitopes were completely exposed on the surface of Jug r 1, suggesting that the predicted regions are able to form linear epitopes.

However, the predicted linear epitopes were not equal to the real ones that caused food allergies. Burks et al. (1997) predicted 13 potential linear epitopes of Ara h 1, the major allergen of peanut, but only

five of them were recognized by the sera of peanut allergic patients. Therefore, overlapping peptide synthesis is a “gold standard” to screen linear epitopes of allergens. In this study, we synthesized 44 overlapping peptides covering the sequence of Jug r 1 rather than the predicted six linear epitopes to avoid omitting possible linear epitopes. Sera from Chinese walnut allergic patients were used to map linear epitopes from the synthetic peptides. Our results showed that AA16-30 and AA125-139 are real linear epitopes of Jug r 1. They were then used to verify the predicted results. We found that linear epitopes predicted by bioinformatics analysis were not consistent with real linear epitopes identified by sera. More accurate and efficient ways for studying the linear epitopes of Jug r 1 are needed.

Based on the pathogenesis of food allergy and the structural features of Jug r 1, we conjecture that the digestion-resistant property of Jug r 1, which is neglected in bioinformatics analysis, is potential direction for studying its linear epitopes. Thus, we used this property to analyze the linear epitopes of Jug r 1 in the present study. Digestion-resistant peptides belonging to Jug r 1 were detected by HPLC-MS/MS after *in vitro* simulated gastrointestinal digestion. In total, we detected five digestion-resistant peptides, including AA19-33, AA40-45, AA54-74, AA96-106, and AA117-137. They were then located in the spatial structure of Jug r 1. The results showed that the four digestion-resistant peptides (AA19-33 is eliminated as the 3D model of Jug r 1 started at residue 35) were completely exposed on the surface of Jug r 1, suggesting that they are possible to be the linear epitopes of Jug r 1. Next, linear epitopes identified by the sera of walnut allergic patients were used to validate whether digestion-resistant peptides are linear epitopes of Jug r 1. Obviously, the digestion-resistant peptides contained all linear epitopes recognized by sera (AA16-30 and AA125-139). We propose that following the strategy of synthesizing digestion-resistant peptides of Jug r 1 to replace 44 overlapping peptides and verifying them using sera from walnut allergic patients, can significantly reduce workload and serum requirement, enhance pertinence, and improve efficiency. Thus, methods based on digestion-resistant properties may be an efficient and targeted strategy for studying the linear epitopes of Jug r 1 compared with conventional methods. Additionally, these methods can offer a reference for further research on the linear epitopes of other food allergens belonging to 2S albumin.

There are some limitations of this study. First, we did not have enough sera samples from Chinese walnut allergic patients. Therefore, AA16-30 and AA125-139 identified by the sera of six walnut allergic patients cannot be representative of the linear epitopes of Jug r 1 in China. Although our results showed that the digestion-resistant property is a promising direction to study the linear epitopes of Jug r 1, more sera samples are needed to validate the feasibility of this method. Second, we performed *in vitro* simulated gastric fluid digestion and *in vitro* simulated intestinal fluid digestion, respectively. In actual situations, digestion is a sequential process. To better simulate the actual situations, the pH of the digestive products of pepsin should be adjusted to 7.5 after digestion for 60 min, and then add the mixture to the simulated intestinal fluid. Further research is needed to improve the experimental procedures of *in vitro* simulated gastrointestinal digestion. Third, although Jug r 1 is stable during digestion, it does not mean that all linear epitopes of Jug r 1 are digestion-resistant. We may miss some important linear epitopes from digestible peptides of Jug r 1, as we only pay attention to digestion-resistant peptides.

4. Conclusion

Although reliable results can be obtained by conventional methods (bioinformatics analysis, overlapping peptide synthesis) when studying linear epitopes of Jug r 1, these methods require a lot of work to synthesize overlapping peptides and are poorly targeted. In contrast, methods based on the digestion-resistant property (*in vitro* simulated gastrointestinal digestion, HPLC-MS/MS) of Jug r 1 are efficient and well-directed and provide convincing results. Our findings showed that

sequence analysis of digestion-resistant peptides may be a promising option for studying linear epitopes of Jug r 1 but needs further verification. In addition, our results contribute to understanding the mechanisms of walnut allergy in China.

CRedit authorship contribution statement

Xiaoya Guo: Conceptualization, Methodology, Data curation, Writing - original draft, Writing - review & editing. **Songsong Jiang:** Investigation, Validation, Writing - original draft, Writing - review & editing. **Xinrui Li:** Validation, Writing - review & editing. **Shuai Yang:** Project administration, Writing - review & editing. **Lei Cheng:** Visualization, Writing - review & editing. **Jinyu Qiu:** Resources, Writing - review & editing. **Huilian Che:** Supervision, Funding acquisition, Writing - review & editing.

Declaration of Competing Interest

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

Acknowledgment

This work was supported by National Major Science and Technology Projects of China (No. 2016zx08011006).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.foodchem.2020.126711>.

References

- Aalberse, R. C. (1997). Food allergens. *Environmental Toxicology and Pharmacology*, 4, 55–60. [https://doi.org/10.1016/S1382-6689\(97\)10042-4](https://doi.org/10.1016/S1382-6689(97)10042-4).
- Arnold, K., Bordoli, L., Kopp, J., & Schwede, T. (2006). The SWISS-MODEL workspace: A web-based environment for protein structure homology modelling. *Bioinformatics*, 22, 195–201. <https://doi.org/10.1093/bioinformatics/bti770>.
- Bertelloni, F., Forzan, M., Turchi, B., Sagona, S., Mazzei, M., Felicioli, A., ... Cerri, D. (2018). A serological survey on swine brucellosis using standard procedures, dot blot, and Western Blot in Finisher Pigs in Central-North Italy. *Veterinary Sciences*, 5, 86. <https://doi.org/10.3390/vetsci5040086>.
- Blythe, M. J., & Flower, D. R. (2010). Benchmarking B cell epitope prediction: Underperformance of existing methods. *Protein Science*, 14, 246–248. <https://doi.org/10.1110/ps.041059505>.
- Bredhorst, R., & David, K. (2001). What establishes a protein as an allergen? *Journal of Chromatography B: Biomedical Sciences and Applications*, 756, 33–40. [https://doi.org/10.1016/S0378-4347\(01\)00069-X](https://doi.org/10.1016/S0378-4347(01)00069-X).
- Breiteneder, H., & Radauer, C. (2004). A classification of plant food allergens. *Journal of Allergy and Clinical Immunology*, 113, 821–830. <https://doi.org/10.1016/j.jaci.2004.01.779>.
- Burks, A. W., Shin, D., Cockrell, G., Stanley, J. S., Helm, R. M., & Bannon, G. A. (1997). Mapping and mutational analysis of the IgE-binding epitopes on Ara h 1, a legume vicilin protein and a major allergen in peanut hypersensitivity. *FEBS Journal*, 245, 334–339. <https://doi.org/10.1111/j.1432-1033.1997.t01-1-00334.x>.
- Downs, M. L., Semic-Jusufagic, A., Simpson, A., Bartra, J., Fernandez-Rivas, M., Rigby, N. M., ... Mills, E. N. C. (2014). Characterization of low molecular weight allergens from english walnut (*Juglans regia*). *Journal of Agricultural and Food Chemistry*, 62, 11767–11775. <https://doi.org/10.1021/jf504672m>.
- Emini, E. A., Hughes, J. V., Perlow, D. S., & Boger, J. (1985). Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *Journal of virology*, 55, 836–839. <https://doi.org/10.1128/jvi.55.3.836-839.1985>.
- World Health Organization. Evaluation of allergenicity of genetically modified foods (2001). <https://www.who.int/foodsafety/publications/gmo-allergenicity/en/>. Accessed 20.03.25.
- Goodman, R. E., & Hefle, S. L. (2005). Gaining perspective on the allergenicity assessment of genetically modified food crops. *Expert Review of Clinical Immunology*, 1, 561–578. <https://doi.org/10.1586/1744666X.1.4.561>.
- He, Y., & Tao, A. (2015). Bioinformatics Methods to Predict Allergen Epitopes. In A. Tao & E. Raz Bioinformatics Methods to Predict Allergen Epitopes (Eds), Allergy Bioinformatics (pp. 223–238). Dordrecht: Springer Science + Business Media B.V.
- Hopp, T. P., & Woods, K. R. (1981). Prediction of protein antigenic determinants from amino acid sequences. *Proceedings of the National Academy of Sciences*, 78, 3824–3828. <https://doi.org/10.1073/pnas.78.6.3824>.

- Jameson, B. A., & Wolf, H. (1988). The antigenic index: A novel algorithm for predicting antigenic determinants. *Computer Applications in the Biosciences Cabios*, 4, 181–186. <https://doi.org/10.1093/bioinformatics/4.1.181>.
- Jennifer, J. S. C., Sydne, J. N., Marc, A. R., Dena, M. B., Margaret, M., Marika, J. S., ... Benjamin, J. H. (2010). Diagnosing and managing common food allergies: A systematic review. *Journal of the American Medical Association*, 303, 1848–1865. <https://doi.org/10.1001/jama.2010.582>.
- Jung, G., & Beck-Sickinger, A. G. (2010). Multiple peptide synthesis methods and their applications. New synthetic methods. *Angewandte Chemie International Edition*, 31, 367–383. <https://doi.org/10.1002/chin.199229317>.
- Karplus, P. A., & Schulz, G. E. (1985). Prediction of chain flexibility in proteins. *The Science of Nature*, 72, 212–213. <https://doi.org/10.1007/BF01195768>.
- Kimber, I., & Dearman, R. J. (2001). Food allergy: What are the issues. *Toxicology Letters*, 120, 165–170. [https://doi.org/10.1016/s0378-4274\(01\)00274-0](https://doi.org/10.1016/s0378-4274(01)00274-0).
- Kyte, J., & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology*, 157, 105–132. [https://doi.org/10.1016/0022-2836\(82\)90515-0](https://doi.org/10.1016/0022-2836(82)90515-0).
- Leung, P. S. C., Chu, K. H., Chow, W. K., Ansari, A., Bandea, C., Kwan, H. S., ... Gershwin, M. E. (1994). Cloning, expression, and primary structure of *Metapenaeus ensis*, tropomyosin, the major heat-stable shrimp allergen. *Journal of Allergy and Clinical Immunology*, 94, 882–890. [https://doi.org/10.1016/0091-6749\(94\)90156-2](https://doi.org/10.1016/0091-6749(94)90156-2).
- Lin, J., Bardina, L., Shreffler, W. G., Andreae, D. A., Ge, Y., Wang, J., ... Sampson, H. A. (2009). Development of a novel peptide microarray for large-scale epitope mapping of food allergens. *Journal of Allergy and Clinical Immunology*, 124, 315–322.e3. <https://doi.org/10.1016/j.jaci.2009.05.024>.
- Lin, X., Yang, H., Chen, H., Wu, J., Liu, Y., & Wei, J. (2014). In Silico Prediction of T and B Cell Epitopes of Der f 25 in *Dermatophagoides farinae*. *International Journal of Plant Genomics*, 2014, 483905. <https://doi.org/10.1155/2014/483905>.
- Liu, C., & Sathe, S. K. (2018). Food Allergen Epitope Mapping. *Journal of Agricultural and Food Chemistry*, 68, 7238–7248. <https://doi.org/10.1021/acs.jafc.8b01967>.
- Liu, Z. Q., Li, X. X., Qiu, S. Q., Yu, Y., Li, M. G., Yang, L. T., ... Yang, P. C. (2017). Vitamin D contributes to mast cell stabilization. *Allergy*, 72, 1184–1192. <https://doi.org/10.1111/all.13110>.
- Mari, A., Scala, E., Palazzo, P., Ridolfi, S., Zennaro, D., & Carabella, G. (2006). Bioinformatics applied to allergy: Allergen databases, from collecting sequence information to data integration. The Allergome platform as a model. *Cellular Immunology*, 244, 97–100. <https://doi.org/10.1016/j.cellimm.2007.02.012>.
- Macdonald, T. T., & Monteleone, G. (2005). Immunity, inflammation, and allergy in the gut. *Science*, 307, 1920–1925. <https://doi.org/10.1126/science.1106442>.
- Moreno, F. J., Mellon, F. A., Wickham, M. S. J., Bottrill, A. R., & Mills, E. N. C. (2005). Stability of the major allergen Brazil nut 2S albumin (Ber e 1) to physiologically relevant in vitro gastrointestinal digestion. *FEBS Journal*, 272, 341–352. <https://doi.org/10.1111/j.1742-4658.2004.04472.x>.
- Pahr, S., Selb, R., Weber, M., Focke-Tejkl, M., Hofer, G., Dordić, A., ... Valenta, R. (2014). Biochemical, biophysical and IgE-epitope characterization of the wheat food allergen, Tri a 37. *PLOS ONE*, 9, e111483. <https://doi.org/10.1371/journal.pone.0111483>.
- Rosenfeld, L., Shreffler, W., Bardina, L., Niggemann, B., Wahn, U., Sampson, H., & Beyer, K. (2012). Walnut allergy in peanut-allergic patients: Significance of sequential epitopes of walnut homologous to linear epitopes of Ara h 1, 2 and 3 in relation to clinical reactivity. *International Archives of Allergy and Immunology*, 157, 238–245. <https://doi.org/10.1159/000327841>.
- Roux, K. H., Teuber, S. S., & Sathe, S. K. (2003). Tree nut allergens. *International Archives of Allergy and Immunology*, 131, 234–244. <https://doi.org/10.1159/000072135>.
- Shanti, K. N., Martin, B. M., Nagpal, S., Metcalfe, D. D., & Rao, P. V. (1993). Identification of tropomyosin as the major shrimp allergen and characterization of its IgE-binding epitopes. *Journal of immunology*, 151, 5354–5363.
- Sigrist, C. J. A., Castro, E., Cerutti, L., Cuche, B. A., Hulo, N., Bridge, A., ... Xenarios, I. (2013). New and continuing developments at PROSITE. *Nucleic Acids Research*, 41, D344–D347. <https://doi.org/10.1093/nar/gks1067>.
- Sun, N., Zhou, C., Zhou, X., Sun, L., & Che, H. L. (2015). Use of a rat basophil leukemia (RBL) cell-based immunological assay for allergen identification, clinical diagnosis of allergy, and identification of anti-allergy agents for use in immunotherapy. *Journal of Immunotoxicology*, 12, 199–205. <https://doi.org/10.3109/1547691X.2014.920063>.
- Toomer, O. T., Do, A., Pereira, M., & Williams, K. (2013). Effect of simulated gastric and intestinal digestion on temporal stability and immunoreactivity of peanut, almond, and pine nut protein allergens. *Journal of Agricultural and Food Chemistry*, 61, 5903–5913. <https://doi.org/10.1021/jf400953q>.
- Teuber, S. S., Dandekar, A. M., Peterson, W. R., & Sellers, C. L. (1998). Cloning and sequencing of a gene encoding a 2S albumin seed storage protein precursor from English walnut (*Juglans regia*), a major food allergen. *Journal of Allergy and Clinical Immunology*, 101, 807–814. [https://doi.org/10.1016/S0091-6749\(98\)70308-2](https://doi.org/10.1016/S0091-6749(98)70308-2).
- Yang, X., & Yu, X. (2009). An introduction to epitope prediction methods and software. *Reviews in Medical Virology*, 19, 77–96. <https://doi.org/10.1002/rmv.602>.

Table 1. The characteristic information of walnut allergic patients

Number	Gender	Age	Symptoms	Other histories of food allergies
1	male	26	nausea, tinnitus	apples, peanuts
2	female	33	stomachache, tinnitus, nausea	peanuts, hazelnuts
3	female	45	stomach cramps, vomiting	eggs
4	male	5	diarrhea, skin rash	peanuts
5	female	7	skin rashes, asthma, tinnitus	hazelnuts
6	male	17	asthma, tinnitus	none

Table 2. Predicted linear epitopes of three immunoinformatic-based computational approaches

Predictive tools	Predicted linear epitopes
DNAStar Protean	28-35, 42-49, 55-62, 65-73, 97-104, 109-121
BPAP	4-12, 48-58, 73-80, 86-99, 121-134
BepiPred 1.0 server	22-49, 58-72, 82-90, 98-129
Final results	28-35, 42-49, 55-62, 65-73, 97-104, 109-121

Table 3. Peptides matched with Jug r 1 in the mixture of Jug r 1 digested by pepsin

Ordinal number	Location	Theoretical molecular weight	Observed molecular weight	Error	Peptides sequence
1	54-65	1477.5797	1477.7182	-0.1385	YYLRQQSRSGGY
2	56-74	2321.3819	2321.0765	0.3054	LRQQSRSGGYDEDNQRQHF
3	57-74	2208.2243	2207.9924	0.2319	RQQSRSGGYDEDNQRQHF
4	59-74	1923.9094	1923.8328	0.0766	QSRSGGYDEDNQRQHF
5	65-74	1351.3372	1352.5501	-1.2129	YDEDNQRQHF
6	96-106	1339.5038	1339.7552	-0.2514	VVRRQQQQQGL
7	117-137	2377.6175	2375.1302	2.4873	SARDLPNECGISSQRCEIRRS

Table 4. Peptides matched with Jug r 1 in the mixture of Jug r 1 digested by pancreatin

Ordinal number	Location	Theoretical molecular weight	Observed molecular weight	Error	Peptides sequence
1	19-33	1762.8896	1821.8356	-58.946	TITTMEIDEDIDNPR
2	22-33	1447.5242	1490.6315	-43.1073	TMEIDEDIDNPR
3	23-33	1346.4203	1346.5892	-0.1689	MEIDEDIDNPR
4	24-33	1215.2242	1215.5487	-0.3245	EIDEDIDNPR
5	25-33	1086.1102	1086.5061	-0.3959	IDEDIDNPR
6	26-33	972.9526	973.4221	-0.4695	DEDIDNPR
7	27-33	857.8652	858.3951	-0.5299	EDIDNPR
8	28-33	728.7512	771.3631	-42.6119	DIDNPR
9	40-45	828.9167	871.4744	-42.6119	REIQR