Germin-like Protein 9-3, a Potential Allergen from a Rice Variety derived from Radiation-Mutation Breeding: An In Silico Study

Custer C. Deocaris[1, 2*], Malona A. Alinsug[1, 3], Alma E. Nacua[4] and Elmer-Rico E. Mojica[5]

[1]Atomic Research Division
Department of Science and Technology – Philippine Nuclear Research Institute (DOST-PNRI), Quezon City, 1100 Philippines
[*]*ccdeocaris@pnri.dost.gov.ph*

[2]Research and Development Management Office
Technological Institute of the Philippines, Quezon City, 1100 Philippines

[3]School of Graduate Studies
Mindanao State University – General Santos, General Santos City, 9500 Philippines

[4]College of Arts and Science, Universidad de Manila
Ermita, Manila 1000 Philippines

[5]Department of Chemistry and Physical Sciences
Pace University, New York, 10029 United States

Date received: February 13, 2021
Revision accepted: August 19, 2021

Abstract

New varieties of crops with a history of safe and widespread use derived from radiation-mutation breeding are generally regarded to be unlikely to produce food safety concerns compared with genetically modified (GM) crops. However, more recent molecular analysis revealed that there might be a greater likelihood that the former may bring about more unintended effects than genetic modification by foreign gene introduction. Thus, in this paper, an allergenicity prediction analysis was carried out on the microarray data of a stable rice mutant derived from γ irradiation-mutation breeding. For the first time, it was revealed that germin-like protein (GLP) 9-3, the transcript with the most significant fold increase in expression in a transcriptome data analyzed, may be a potential allergen in radiation-mutant rice. An immunoinformatic analysis of GLP 9-3 disclosed several linear and conformational (discontinuous) B-cell epitopes and MHCII-binders. These epitopes were located primarily on the random coil and surface-exposed regions based on inspection of the protein's secondary structures. Additionally, the GLP 9-3 epitopes were predicted to be reactive in more than 50% of the population in top rice-producing and -consuming countries based on human leukocyte antigen (HLA) haplotypes. Since this is the first report on the up-regulation of GLP 9-3 allergen in mutant rice, it is recommended to investigate this molecule for further biochemical and clinical analyses.

***Keywords**:* food allergens, radiation mutation breeding, unintended effects, food safety, transcriptome

1. Introduction

Mutation breeding is a highly successful tool in the global efforts to produce new plant varieties to help feed an ever-increasing human population. As mutations occur spontaneously, they can be induced by either physical or chemical treatments. The induction of mutations has been used to enhance the yield, fortify the nutritional quality, and improve the environmental and climate adaptability of the world's most important crops such as wheat, rice and corn. For crop improvement, the physical mutagens usually employed are X-rays and gamma rays. In contrast, ethyl methanesulfonate, diethyl sulfate, ethylenimine, sodium azide and colchicine are the most frequently used chemical mutagens.

Recently, the heavy-ion beam has also emerged as an efficient and robust approach for inducing mutations. This high-linear energy transfer (LET) physical mutagen, compared with the low-LET X-rays and gamma-rays, can result in a higher mutation rate, broader mutation spectrum and shorter breeding cycle (Hu *et al*., 2017). In the more than 80-year-old history of plant breeding based on induced mutations, there are many registered crop variants with new and valuable horticultural characters registered in the Joint International Atomic Energy Agency-Food and Agriculture Organization (IAEA-FAO) Mutant Variety Database. As of 20 January 2021, among the physical mutagens, gamma-rays are more widely employed than X-rays with 1,693 and 569 improved mutants, respectively. The global leading radiation mutant producers are China (817), Japan (479) and India (341). The Philippines, so far, has 20 registered radiation mutants. The mutant rice variety BPI-121-407 was the first officially approved Filipino radiation mutant variety that received certification in 1971. It was developed by irradiation of seeds with gamma rays and mixed neutrons. The main improved attributes of mutant variety are early maturity, very short stem, stiff-strawed, high tillering, resistance to diseases and moderate resistance to bacterial leaf blight (IAEA/FAO, n.d.).

While radiation mutation breeding has a history of enjoying strong public acceptance, in contrast, the use of transgenic technologies for varietal improvement is challenged with enormous controversy regarding the food safety of the resulting products (Matsaunyane and Dubery, 2018). As a result, the production of genetically modified (GM) crops has led to increased investigations within government regulatory boards regarding research to assess their food and environmental safety (Ekici and Sancak, 2011). Many

countries regulate the consumption of genetically modified organisms (GMOs) at a specific percentage of the content and would require labeling the food packages containing GMOs. In the European Union, for example, labeling of food and feed is required where the level of approved GMO exceeds 0.9% of unintentional adventitious presence. For non-approved GMOs, on the other hand, the threshold is 'zero,' requiring cargoes of non-approved GMOs to be returned to the port of origin or even destroyed (Davison, 2010). While there are several benefits ascribed to GM crops, in addition to the strict governmental regulations, many counter the product with issues that include potential toxicity, the assumption that the products may contain allergens and the possible development of antibiotic resistance from the utilization of GM products (Key *et al*., 2008). Thus, the FAO and the European Food Safety Authority (EFSA) recommendations have also called for targeted approaches to evaluate macro-, micro- and anti-nutrients, toxins, allergens and secondary metabolites for GM crops (Kuiper *et al*., 2001).

Some molecular profiling methods have also been proposed to increase the chances of detecting unintended effects. The use of non-targeted global profiling (e.g., microarray technology) is an effective means of identifying unintended effects in several transgenic plant systems (Baudo *et al*., 2006; Batista *et al*., 2008; Abdeen *et al*., 2010; Barros *et al*., 2010; Montero *et al*., 2011; Ricroch *et al*., 2011; Ricroch, 2013; Schnell *et al*., 2015; Ko *et al*., 2018; Fu *et al*., 2019; Long *et al*., 2020). The potential toxicity of GM crops is believed to arise from these unintended pleiotropic effects of the transgene, containing the coding region and regulatory elements, on the host plant genome (Filipecki and Malepszy, 2006). Since integrating a transgene in the host genome is usually a random process, additional unintended effects may also occur from host gene disruption and deoxyribonucleic acid (DNA) sequence rearrangements at the insertion site or proximal host sequences (Forsbach *et al*., 2003; Latham *et al*., 2006).

Evidence supports that radiation mutation breeding could lead to even more unintended effects than transgenesis. In a study by Batista *et al*. (2008), the observed alteration was more extensive in radiation-mutagenized than in GM plants in all the cases studied. For example, 11,267 genes showed differential expression in the non-stable radiation mutant rice line, whereas only 2,318 genes were detected in the non-stable transgenic line. The group recommended that "safety assessment of improved plant varieties should be carried out on a case-by-case basis and not simply restricted to foods obtained through genetic engineering" (Batista *et al*., 2008, p. 3644). Since it is an

accepted scientific opinion that new GM crops should be screened for potential food allergies, it may, thus, be prudent to search for potential allergens in radiation-mutant crops since the latter can have a higher potential for unintended effects.

Food allergy represents a significant public health problem affecting approximately 2-4% of the adult population and 8-9% of children diagnosed. The prevalence rate for self-reported food allergy is several times higher (Ververis *et al*., 2020). This study characterized the allergenic potential of germin-like protein (GLP) 9-3, one of the major allergens that the authors previously discovered up-regulated in a radiation mutant rice variety. Germins and the related GLPs are glycoproteins expressed in many plants in response to biotic and abiotic stress. These were found to represent a new class of food allergens (Jensen-Jarolim *et al*., 2002).

2. Methodology

2.1 Screening of Allergens from Microarray Data

Data mining for allergens from microarray data will be reported in a separate paper. Briefly, the gene expression profile GSE12069, which was last updated on December 27, 2017, was downloaded from the Gene Expression Omnibus (GEO) database (Barrett *et al*., 2013). It contains 48,564 transcripts of the genetically stable dwarf-mutant rice cultivar Estrela A (*Oryza sativa L. ssp. Japonica*). It had already gone over 10 generations of self-pollination. The variety was developed from Estacâo Agrŏnomica Nacional (Oeiras, Portugal) in 1988 by gamma-irradiation (Batista *et al*., 2008). A list of differentially expressed genes (DEGs) that are significantly up-regulated compared with the control mother plant ($p < 0.01$) was collected.

After determining the allergenicity status and expression levels of GLP 9-3 from the allergenicity screening, all non-redundant sequences of GLPs from rice (cv. Japonica) was obtained from the GenBank (National Center for Biotechnology Information [NCBI], 1988). After obtaining the FASTA protein sequences, allergens were predicted using AlgPred (Lafarga *et al*., 2016), Structural Database of Allergenic Proteins (SDAP) (Ivanciuc *et al*., 2003) and AllergenFP (Dimitrov *et al*., 2013) servers following on the current

FAO/World Health Organization (WHO) Codex Alimentarius guidelines (FAO/WHO, 2001).

2.2 Phylogenetic Analysis

Nucleotide sequences of GLP3 in FASTA formats were aligned based on their conserved regions employing Multiple Alignment using Fast Fourier Transform (MAFFT). MAFFT FASTA-formatted sequences were then keyed in for phylogenetic analyses using the Molecular Evolutionary Genetics Analysis X program (Kumar *et al*., 2018). Tree reconstruction was inferred using the Neighbor-Joining method with a Poisson correction model and a bootstrap test of 1,000 replicates. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test was shown next to the branches. Branch lengths were reflected as the same units as those of the evolutionary distances used to infer the phylogenetic tree.

2.3 Physicochemical and Secondary Structural Characterization

ProtParam tool was used to assess the physicochemical properties of the protein sequences. The parameters analyzed included molecular weight, theoretical pI, amino acid composition, the total number of negatively charged residues (Asp + Glu), instability index, aliphatic index, the total number of positively charged residues (Arg + Lys) and the grand average of hydropathicity (GRAVY) (Gasteiger *et al*., 2005). The secondary structure, relative surface accessibility and disordered region were predicted using NetSurfP version 2.0 (Klausen *et al*., 2019).

2.4 B-cell Epitope Mapping

Two types of B-cell epitopes were predicted: linear and discontinuous. For the linear epitopes, Bepipred Linear Epitope Prediction (BepiPred) version 2.0 tool was utilized. BepiPred calculates the probability that a given antigen residue is part of an epitope using a Random Forest Regression algorithm (Jespersen *et al*., 2017) on a window of nine residues. The string of amino acids above the threshold value of 0.5 was considered statistically significant. Since it has been estimated that > 90% of B-cell epitopes are regarded as discontinuous, DiscoTope was employed to predict the surface epitopes accounting for the 3D model of the query protein (Kringelum *et al*., 2012). Residues with threshold value above -7.7 was regarded as potential epitope.

2.5 T-cell Epitope Mapping

Fifteen-mer T-cell epitopes restricted to HLA-DRB1*03:01, HLA-DRB1*07:01, HLA-DRB1*15:01, HLA-DRB3*01:01, HLA-DRB3*02:02, HLA-DRB4*01:01 and HLA-DRB5*01:01 were predicted using the online prediction applications in Immune Epitope Database (IEDB) (Wang *et al*., 2008, 2010). After a consensus analysis, the median of the percentile ranks less than 2.0 was considered.

2.6 Homology Modeling and Structural Validation

The 3D protein structure was modeled on the SWISS-MODEL workspace using the alignment mode with energy minimization value (Waterhouse *et al*., 2018) and visualized using PyMol (Yuan *et al*., 2016). The stereochemical property of the models was checked by PROCHECK (Laskowski *et al*., 1993).

2.7 Analysis of Population Coverage of Epitopes

To understand the degree of the allergen coverage in the human population, epitope sequences were submitted to the population coverage analysis tool of IEDB by maintaining the default analysis parameters (Bui *et al*., 2006).

3. Results and Discussion

3.1 GLP 9-3 as an Allergen in Rice

A significant finding from the allergenicity screening of the transcriptome of a radiation-mutant of *O. sativa* from the paper of Batista *et al*. (2008) was the up-regulation of GLP 9-3 (XP_015612248.1). The transcript had an increased gene expression of 5.8 folds in the stable mutant line compared with the control plant. It also had the highest increase in transcript level among the putative allergenic genes identified. From the five algorithms employed, SDAP, AlgPred, AllergenFP and Allermatch gave positive results. It was also noted that GLP 8-7 was found up-regulated in GMO rice (Deocaris *et al*., 2020).

GLPs are glycoproteins expressed in many plants in response to biotic and abiotic stress. These proteins were first discovered in wheat seeds as specific markers of germination. Later, these were found in other monocotyledons,

dicotyledons, gymnosperms and moss (Dunwell *et al*., 2008). The germin family belongs to the functionally diverse cupin superfamily, codes for two exons and contains the cupin motif (PF00190) at its C-terminus (Barman and Banerjee, 2015). The diverse biochemical activities and functions ascribed to GLPs are reflected in the way this class of proteins is referred to (i.e., oxalate oxidase-like proteins, nectarines, rhicadhesin-like receptors and ADP glucose pyrophosphatase/phosphodiesterases [AGPPase]) (Davidson *et al*., 2009).

3.2 Phylogenetic Tree and Clustering of Allergenic GLPs

To identify other GLP members in rice, the investigators queried the genome of *O. sativa* (taxid: 4530) in the GenBank (NCBI, 1988) using "germin-like protein" as search term. After removing redundant sequences and hypothetical proteins, 41 GLP family member proteins in 10 chromosomes (Chr. 1, 2, 3, 4, 5, 7, 8, 9, 11 and 12) remained. To further validate the GLP family members, the researchers performed Simple Modular Architecture Research Tool (SMART) analysis of the protein sequences, which detected the Cupin_1 (PF00190) domain. Among the remaining 41 members, there were two GLP 5-1 proteins with different amino acid sequences which were both located on chromosome 5. The investigators operationally designated these similarly named proteins as GLP 5-1.1 (accession Q688L5; 230 aa) and GLP 5-1.2 (accession Q6I544; 221 aa).

To gain insights into the relationship of the 41 GLPs in terms of allergenic potential in the radiation-mutant rice, hierarchical clustering analysis of the protein sequences was done. Phylogenetic analysis of 41 GLPs was conducted through Neighbor-Joining method using the Molecular Evolution Genetic Analysis 7 (MEGA 7) software (Figure 1). It was found that the 41 GLPs were distributed across four clusters. Cluster 1 was predominantly composed of GLP subfamilies 2, 8 and 12; Cluster 2, subfamilies 1, 3, 4, 5; Cluster 3, subfamilies 1, 5 and 6; and Cluster 4, subfamilies 2, 3, 8, 9 and 11. Cluster 1 had the higher number of potentially allergenic GLP members (10/19 = 52.6%). This GLP cluster was followed by Cluster 4 (3/8 = 37.5%), Cluster 2 (2/6 = 33.3%) and Cluster 3 (1/6 = 12.5%). While the significant gene expression differences were most prominent in Cluster 1, the three GLPs (GLP 8-11, GLP 8-2 and GLP 8-12) were all downregulated in the mutant variety compared with the wildtype. It is noteworthy that in Cluster 4, two out of three members of the GLP subfamily 9, GLP 9-1 and GLP 9-3, were both up-regulated in the mutant plant.

The high-energy photons of gamma radiation penetrate and interact with living tissues. While it is used to create mutations, the energy decreases

growth rate and reproduction capacity. An adaptive strategy of plants under such stress includes expressing stress-associated transcription factors (TFs). These TFs regulate many pathogenesis-related or signaling genes after binding to their respective promoters (Santino *et al*., 2013).

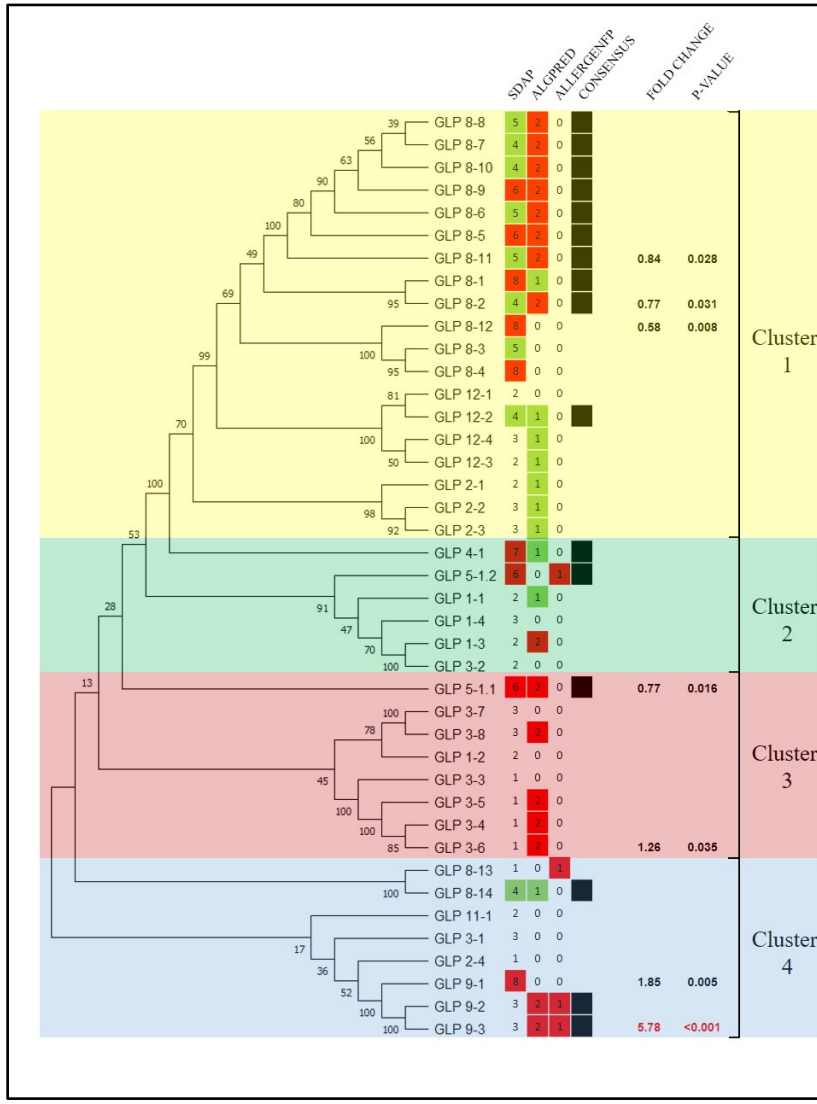

Figure 1. Hierarchical clustering analysis of rice germin-like proteins displaying four distinct clusters

In Figure 1, juxtaposed to the tree were the results of allergenicity screening sig SDAP, AlgPred and AllergenFP servers. The numbers indicate potential allergenic sequences identified. Red and green shaded numbers relate to 'strong' and 'moderate' allergenicity potentials, respectively. A 'consensus' assignment of allergenicity is provided for each GLP member (black shade) when two of the three servers indicated either strong or moderate allergenicity. The fold-change and p-values are shown based on the microarray data by Batista *et al.* (2008). Only the P-values < 0.05 are shown.

Interestingly, the pattern of up-regulation of GLP 9-1 and GLP 9-3 transcripts with radiation-mutation breeding is associated with the nature of the transcription factors binding (TFb) sites. In a bioinformatics work by Das *et al*. (2019), the promoter regions of rice GLP 9-1 (22697583–22696084) and GLP 9-3 (22702365–22700866) are the most related in the GLP family in terms of the number and types of stress-associated TFb sites. The promoter of GLP 9-1 contains four NAM/ATAF1/CUC2 (NAC), two N-terminal WRKY domain-containing TF (WRKY), three basic helix-loop-helix (bHLH), five basic leucine zipper (bZIP), five myeloblastosis viral oncogene homolog TF (MYB) and 93 APETALA 2/ethylene-responsive element binding factor (AP2/ERF) binding sites. The promoter of GLP 9-3 comprises three NAC, two WRKY, five bHLH, three bZIP, seven MYB and 70 AP2/ERF TFb sites.

When plants are selected to acquire desired traits during crop improvement, regardless of whether the genetic modification was via radiation-mutation or transgenesis, the process is accompanied by stimulation in transcript levels of untargeted stress-related genes. Apparently, "the stressing event is memorized along several generations" (Batista *et al*., 2008, p. 3641). This transgenerational memory of stress from radiation mutation or by deliberate genetic modification may likely be attributed to epigenetic mechanisms as reviewed by others. Plants utilize transgenerational epigenetic stress-adaption, where the environmental conditions experienced by the parental generation, is inherited by the offspring; this stress-memory allows future generation to respond either faster or stronger to a reoccurring stimulus (Weinhold, 2018). From the perspective of this study, it would make much sense that those genes triggered by the plant's life history during radiation mutation breeding were a stress protein with allergenic potential – as with the case of these two GLPs.

3.3 Physicochemical Properties of GLP 9-3

GLPs represent a new class of food allergens (Jensen-Jarolim *et al*., 2002). The family includes ubiquitous glycoproteins that are similar to several plant

food allergens including the lemon peel (flavedo) germin-like allergen (Cit l 1), hazelnut allergen (Cor a 9), and peanut allergen (Ara h 3) (Hirano *et al*., 2016). In an immunoblotting assay, 24 out of 82 tested sera (29.26%) from allergic patients showed IgE-binding to germins (Jensen-Jarolim *et al*., 2002). Likewise, the 24 kDa GLP 8-7 was recently reported to be an up-regulated potential allergen from GMO rice (Deocaris *et al*., 2020). According to SDAP, GLP 9-3 bears significant similarity to soybean allergen Gly m Bd 28K, which causes atopic dermatitis in soybean-sensitive patients (Hiemori *et al*., 2000). Gly m Bd 28K comprises the significant allergens of soybean, which include the major seed storage proteins 11S glycinin (Gly m 6), 7S β-conglycinin (Gly m 5) and the oil body-associated Gly m Bd 30K (Kern *et al*., 2019).

The physicochemical properties of GLP 9-3 were compared with the similar allergens identified by the allergen analytical servers. The physiochemical characteristics of the proteins, as predicted using ProtParam, are presented in Table 1.

Table 1. Physicochemical properties of GLP 9-3 and similar allergens from *O. sativa* (rice), *Glycine max* (soybean) and *Arabidopsis thaliana* (thale cress)

	GLP 9-3	Gly m Bd 28K	GLP 3-3	GLP 8-2
Species	*O. sativa*	*G. max*	*A. thaliana*	*O. sativa*
Amino acid residues	214	476	211	221
Molecular weight	22,456.06	52,944.36	21,836.17	23,712.37
Total number of atoms	3175	7425	3100	3368
Theoretical pI	5.45	5.73	6.26	6.40
Aliphatic index	92.10	80.95	99.43	101.95
Instability index	34.16 (stable)	46.57 (unstable)	32.85 (stable)	21.97 (stable)
Grand average of hydropathicity (GRAVY)	0.382	-0.267	0.404	0.220

As expected, GLP 9-3 possessed similar features to the two other GLPs. Based on computed pI values, GLP 9-3 and Gly m Bd 28K were slightly acidic compared with the two other GLPs. The stability of the proteins was investigated by analyzing the instability, aliphatic and GRAVY indices of these three proteins. There were more positively charged residues compared with negatively charged residues for GLP 9-3, Gly m Bd 28K and GLP 8-2. In the case of GLP subfamily 3 member 3, there was slightly more positively

charged residues observed. The instability index values were less than 40 for the GLPs compared with Gly m Bd 28K; therefore, it is likely that GLPs are more stable allergens. The aliphatic index (AI) is defined as the relative volume occupied by aliphatic side chains (alanine, valine, isoleucine and leucine). The AI values revealed that all the proteins were likely to be thermostable. The GLPs' GRAVY scores were found positive with a range from 0.220 to 0.404 indicating that all three proteins were hydrophobic except for Gly m Bd 28K, whose GRAVY score was -0.267. GLPs have been reported to be resistant to proteases, extreme pH, heat and sodium dodecyl sulfate (Zimmermann *et al*., 2006). These properties are all indicative of the allergenic nature of the protein.

3.4 Structural and Immunologic Features of GLP 9-3

Secondary and tertiary protein structures are essential for the prediction of epitopes. Secondary structure prediction with NetSurfP ver. 2.0 identified eight α-helices and 13 β-sheets (Figure 2a). The present results revealed that the proportion of random coils, α-helices and β-sheets accounted for 55.1, 15.4 and 29.5% of the secondary structure, respectively. Of interest in characterizing epitopes are the hydrophilic random coil regions usually located on the protein surface. Such features are essential in serving as binding ligands, and thus, have a high rate of having antigenic epitopes found therein (Sikic *et al*., 2010; Li *et al*., 2013; Tahmoorespur *et al*., 2017).

In Figure 2a, the amino acid sequence and location of epitopes in relation to the secondary structural elements are indicated. The structural features are displayed, indicating relative surface accessibility (RSA), secondary structure (SS) and disordered region (DR). For RSA, the red and blue lines show exposed and buried residues at a 25% threshold, respectively. The α-helices (yellow spirals), β-sheets (solid blue arrows) and random coils (purple lines) are illustrated on the SS panel. The thickness of the gray line displays the probability of disordered states in the DR. The denoted epitope clusters (Clusters 1, 2 and 3) were determined visually based on their proximal positions within the 3D structure of the protein. Figure 2b shows the diagrams of scores for linear and discontinuous B-cell epitopes and percentile ranks for MHC II epitopes. The x-axis exhibits amino acid residues' position in the sequence while the y-axis shows the corresponding score for each amino acid residue. Only the data above the threshold values are shown. Figure 2c projects a 3D structure of GLP 9-3 derived using SWISS-Model and rendered with PyMol. Epitopes are labeled with different colors: B-cell epitopes (red), MHCII-epitopes (blue) and overlapping B-cell and MHCII-epitopes (green).

The IEDB suite identified 57 linear 9-mer B-cell epitopes and 25 15-mer MHCII-binding peptides (or TH-cell epitopes) based on the high stringent threshold used in this study (Figure 2b).

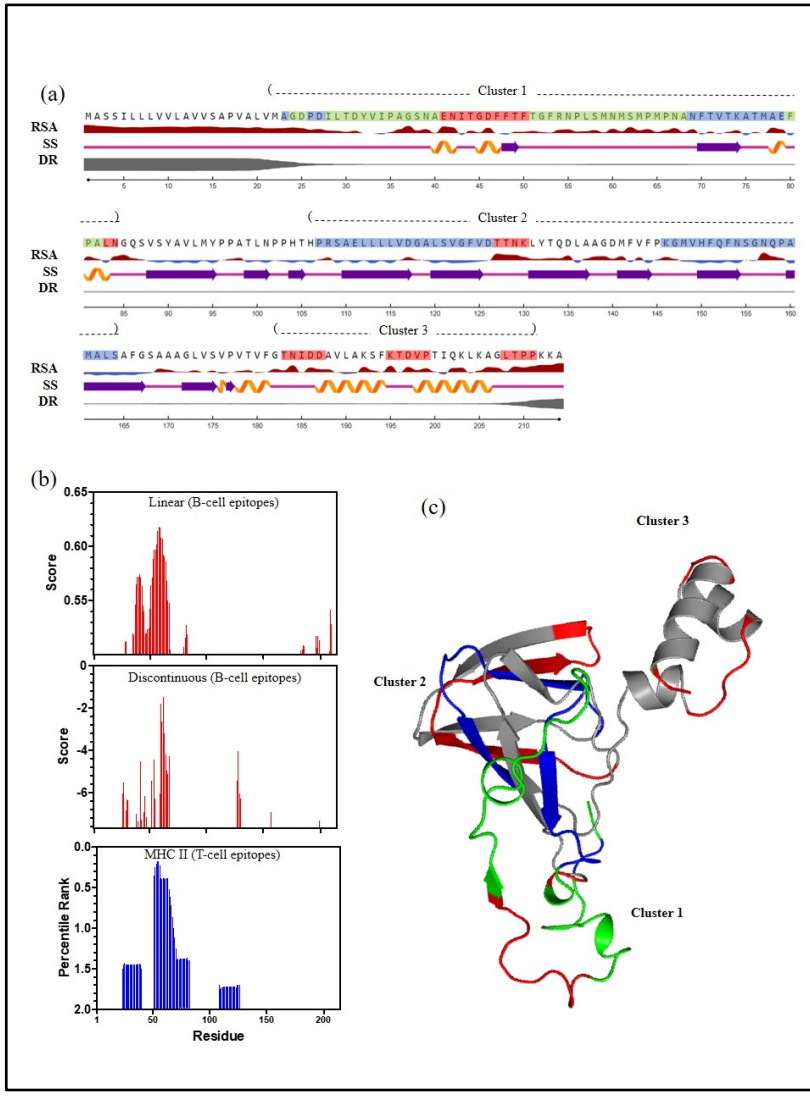

Figure 2. Structural features and mapping of predicted B-cell and MHC II-epitopes on GLP 9-3

The threshold values were set to increase the specificity of the present study's epitope analysis. Additionally, 30 residues were associated with discontinuous epitopes. Thirty-six residues from the predicted epitopes overlapped in the domains of B-cell or TH-cell epitopes. Most of the B-cell epitopes (marked red) reside in random coil and exposed structures. TH-cell epitopes (marked blue), on the other hand, appeared to be more distributed across the protein regardless of the secondary structure. This observation may be consistent with the process of MHC class II loading. During this process, proteins are endocytosed and digested in lysosomes; individual peptide fragments, whatever the secondary structure, are loaded onto the groove of an MHC class II molecules (Blum *et al*., 2013).

Secondary and tertiary protein structures are essential for the prediction of epitopes. While it is easy to determine the secondary structural propensity from the amino acid sequence, the tertiary structure of GLP 9-3 is yet to be empirically solved. Thus, a tentative model of this putative allergen by homology modeling using SWISS-MODEL is shown in Figure 2c (note: only one subunit of the hexamer is displayed). Based on comparative modeling, from 50 templates obtained, the closest template with which the model was built was from the 1.7 Å resolution structure of oxalate oxidase 1 (EC 1.2.3.4) from barley, *Hordeum vulgare* (PDB ID: 2et7) (Opaleye *et al*., 2006), which has a sequence identity of 34.62% and a low E-value of 3e-33. Analysis of the models indicated an acceptable probability of confirmation with 90.6, 8.8, 0.6 and 0% of the residues of GLP 9-3 in the favorable, allowed, generally allowed and disallowed regions of the Ramachandran plot, respectively. For 2et7 template, 93.4% residues were within the most favored regions, 10.6% residues in the additional allowed region, 0% residues in the generously allowed regions and 0% residues in the disallowed region. It is generally accepted that if 90% of residues are present in the allowed regions of a Ramachandran plot, the model is reliable. Hence, based on these validations, the homology model was adopted in this study (Figure 2c).

Based on the proximity of the epitopes along the protein, it was evident that GLP 9-3 contained three epitope clusters. The investigators failed to find any epitope in the disordered regions partitioned at the N- and C-termini of the protein. Of note, intrinsically disordered proteins have high flexibility and lack stable secondary and tertiary structures. These physical characteristics are considered to allow the structure to adopt multiple interactions at the interface of protein-protein interaction networks including antigen-antibody interactions (Sormanni *et al*., 2015; Fusco *et al*., 2018).

3.5 Epitope Coverage of GLP 9-3 in Top Rice-producing Countries

To better understand this potentially allergenic protein's impact on the consumers, a population coverage analysis of GLP 9-3 epitopes was conducted. The countries chosen were among the top producers/consumers of rice globally (Ito, 2019). Myanmar, Bangladesh and Lao PDR were not included in the assessment because their comprehensive HLA data were not available in IEDB. The epitopes have a wide range of projected population coverage ranging from 19.22% (Philippines) to 76.1% (South Korea) and with an average of 51.16% in the 11 countries queried. The population coverage analysis of the selected epitopes is tabulated in Table 2.

Table 2. Population coverage from top rice-consuming countries by the GLP 9-3 epitopes

Country	Coverage (%)[a]	Average Hit[b]	PC90[c]
Brazil	39.46	1.31	0.50
China	50.03	1.71	0.60
Guinea-Bissau	57.36	2.01	0.70
India	66.19	2.39	0.89
Indonesia	41.60	1.37	0.51
Japan	64.54	2.32	0.85
South Korea	76.10	2.95	1.26
Philippines	19.22	0.60	0.37
Taiwan	53.05	1.82	0.64
Thailand	59.13	2.09	0.73
Vietnam	49.96	1.70	0.60
Average	51.16	1.79	0.68
Standard deviation	14.65	0.60	0.23

[a] – projected population coverage; [b] – average number of epitope hits/HLA combinations recognized by the population; [c] – minimum number of epitope hits/HLA combinations recognized by 90% of the population

Given that there was an average of 0.68 epitopes hit in 90% of the 11 populations analyzed, it prompts the question: why are adverse immune reactions to rice rare despite these seemingly high immunogenicity scores? While it is impossible to envisage the complete avoidance of the "offending" rice proteins by allergic individuals, there is no study about the safe consumption levels of these allergens that can provide a clear answer. One would also argue that oral tolerance may play a big role since rice cultivation is part of national culture and almost all people at all life stages, including infants, in these countries are "rice-eaters." Consistent with the dual-allergen exposure hypothesis, early life consumption of food protein is known to induce oral tolerance and prevent the onset of IgE sensitization at later life (Du

Toit *et al*., 2018). This hypothesis was posited after observing the strong protective association of dietary exposure of infants against the risk of manifesting peanut allergy in children (Fox *et al*., 2009). The present investigators' conjecture on the rarity of rice allergies is also consistent with the observation that in countries where infant-appropriate peanut snacks are available, peanut allergy was also found to be rare (Burks *et al*., 2008; Du Toit *et al*., 2008). Similar relationships between early infant and maternal exposure and reduction of sensitization rates have been published for milk, egg and fish allergens (López-Expósito *et al*., 2009; Katz *et al*., 2010; Koplin *et al*., 2010; Herman and Ladics, 2011; Mermiri *et al*., 2017). However, despite the presumed widespread role of oral tolerance, a risk-benefit analysis for new crops cannot be made based on this phenomenon's generalizability as allergic reactions in individuals are on a case-by-case basis, and that many other factors cannot be controlled in the context of gene-environment interaction.

4. Conclusion and Recommendation

An allergenicity prediction analysis carried out on the transcriptome data of a stable rice mutant derived from γ irradiation-mutation breeding revealed GLP 9-3, the transcript with the most significant fold increase in expression, as a potential allergen in radiation-mutant rice. An immunoinformatic analysis of GLP 9-3 showed several linear and conformational (discontinuous) B-cell epitopes and MHCII-binders located mostly in the random coil and surface-exposed regions based on inspection of the protein's secondary structures. Additionally, the GLP 9-3 epitopes were predicted to be reactive in more than 50% of the population in top rice-producing and -consuming countries based on HLA haplotypes. Since this is the first time that GLP 9-3 has been reported to be a potential allergen in rice, it is recommended that this molecule be investigated for further biochemical and clinical analyses.

5. References

Abdeen, A., Schnell, J., & Miki, B. (2010). Transcriptome analysis reveals absence of unintended effects in drought-tolerant transgenic plants overexpressing the transcription factor ABF3. BMC Genomics, 11(1), 69. https://doi.org/10.1186/1471-2164-11-69

Barman, A.R., & Banerjee, J. (2015). Versatility of germin-like proteins in their sequences, expressions, and functions. Functional and Integrative Genomics, 15(5), 533-548. https://doi.org/10.1007/s10142-015-0454-z.

Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Holko, M., Yefanov, A., Lee, H., Zhang, N., Robertson, C.L., Serova, N., Davis, S., & Soboleva, A. (2013). NCBI GEO: Archive for functional genomics data sets-update. Nucleic Acids Research, 41, D991-5.

Barros, E., Lezar, S., Anttonen, M.J., Van Dijk, J. P., Röhlig, R.M., Kok, E.J., & Engel, K.H. (2010). Comparison of two GM maize varieties with a near-isogenic non-GM variety using transcriptomics, proteomics and metabolomics. Plant Biotechnology Journal, 8(4), 436-451. https://doi.org/10.1111/j.1467-7652.2009.00487.x

Batista, R., Saibo, N., Lourenco, T., & Oliveira, M.M. (2008). Microarray analyses reveal that plant mutagenesis may induce more transcriptomic changes than transgene insertion. Proceedings of the National Academy of Science USA, 105(9), 3640-3645. https://doi.org/10.1073/pnas.0707881105

Baudo, M.M., Lyons, R., Powers, S., Pastori, G.M., Edwards, K.J., Holdsworth, M.J., & Shewry, P.R. (2006). Transgenesis has less impact on the transcriptome of wheat grain than conventional breeding. Plant Biotechnology Journal, 4(4), 369-380. https://doi.org/10.1111/j.1467-7652.2006.00193.x

Blum, J.S., Wearsch, P.A., & Cresswell, P. (2013). Pathways of antigen processing. Annual Review of Immunology, 31, 443-473. https://doi.org/10.1146/annurev-immunol-032712-095910

Bui, H.-H., Sidney, J., Dinh, K., Southwood, S., Newman, M.J., & Sette, A. (2006). Predicting population coverage of T-cell epitope-based diagnostics and vaccines. BMC Bioinformatics, 7(1), 1-5. https://doi.org/10.1186/1471-2105-7-153

Burks, A.W., Laubach, S., & Jones, S.M. (2008). Oral tolerance, food allergy, and immunotherapy: Implications for future treatment. Journal of Allergy and Clinical Immunology, 121(6), 1344-1350. https://doi.org/10.1016/j.jaci.2008.02.037

Das, A., Pramanik, K., Sharma, R., Gantait, S., & Banerjee, J. (2019). In-silico study of biotic and abiotic stress-related transcription factor binding sites in the promoter regions of rice germin-like protein genes. PLoS One, 14(2), e0211887. https://doi.org/10.1371/journal.pone.0211887

Davidson, R.M., Reeves, P.A., Manosalva, P.M., & Leach, J.E. (2009). Germins: A diverse protein family important for crop improvement. Plant Science, 177(6), 499-510. https://doi.org/10.1016/j.plantsci.2009.08.012

Davison, J. (2010). GM plants: Science, politics and EC regulations. Plant Science, 178(2), 94-98. https://doi.org/10.1016/j.plantsci.2009.12.005

Deocaris, C.C., Rumbaoa, R.G., Gavarra, A.M., & Alinsug, M.V. (2020). A preliminary analysis of potential allergens in a GMO rice: A bioinformatics approach.

Open Journal of Bioinformatics and Biostatistics, 4(1), 12-16. https://doi.org/10.17352/ojbb.000007

Dimitrov, I., Naneva, L., Doytchinova, I., & Bangov, I. (2013). AllergenFP: Allergenicity prediction by descriptor fingerprints. Bioinformatics, 30(6), 846-851. https://doi.org/10.1093/bioinformatics/btt619

Du Toit, G., Katz, Y., Sasieni, P., Mesher, D., Maleki, S.J., Fisher, H.R., Fox, A.T., Turcanu, V., Amir, T., Zadik-Mnuhin, G., & Cohen, A. (2008). Early consumption of peanuts in infancy is associated with a low prevalence of peanut allergy. Journal of Allergy and Clinical Immunology, 122(5), 984-991. https://doi.org/10.1016/j.jaci.2008.08.039

Du Toit, G., Sampson, H.A., Plaut, M., Burks, A.W., Akdis, C.A., & Lack, G. (2018). Food allergy: Update on prevention and tolerance. Journal of Allergy and Clinical Immunology, 141(1), 30-40. https://doi.org/10.1016/j.jaci.2017.11.010

Dunwell, J.M., Gibbings, J.G., Mahmood, T., & Saqlan Naqvi, S. (2008). Germin and germin-like proteins: Evolution, structure, and function. Critical Reviews in Plant Sciences, 27(5), 342-375. https://doi.org/10.1080/07352680802333938

Ekici, K., & Sancak, Y.C. (2011). A perspective on genetically modified food crops. African Journal of Agricultural Research, 6(7), 1639-1642.

Filipecki, M., & Malepszy, S. (2006). Unintended consequences of plant transformation: A molecular insight. Journal of Applied Genetics, 47(4), 277-286. https://doi.org/10.1007/bf03194637

Food and Agriculture Organization/World Health Organization (FAO/WHO). (2001). Report of a joint FAO/WHO expert consultation on allergenicity of foods derived from biotechnology. Rome, Italy

Forsbach, A., Schubert, D., Lechtenberg, B., Gils, M., & Schmidt, R. (2003). A comprehensive characterization of single-copy T-DNA insertions in the *Arabidopsis thaliana* genome. Plant Molecular Biology, 52(1), 161-176. https://doi.org/10.1023/A:1023929630687

Fox, A.T., Sasieni, P., du Toit, G., Syed, H., & Lack, G. (2009). Household peanut consumption as a risk factor for the development of peanut allergy. Journal of Allergy and Clinical Immunology, 123, 417-423. https://doi.org/10.1016/j.jaci.2008.12.014

Fu, W., Wang, C., Xu, W., Zhu, P., Lu, Y., Wei, S., Wu, X., Wu, Y., Zhao, Y., & Zhu, S. (2019). Unintended effects of transgenic rice revealed by transcriptome and metabolism. GM Crops and Food, 10(1), 20-34. https://doi.org/10.1080/21645698.2019.1598215

Fusco, G., Sanz-Hernandez, M., Ruggeri, F.S., Vendruscolo, M., Dobson, C.M., & De Simone, A. (2018). Molecular determinants of the interaction of EGCG with ordered and disordered proteins. Biopolymers, 109(10), e23117. https://doi.org/10.1002/bip.23117

Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M.R., Appel, R.D., & Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. In Walker J. M. (Ed.), The proteomics protocols handbook (pp. 571-607). New York, NY: Humana Press.

Herman, R.A., & Ladics, G.S. (2011). Endogenous allergen upregulation: Transgenic vs. traditionally bred crops. Food and Chemical Toxicology, 49(10), 2667-2669. https://doi.org/10.1016/j.fct.2011.07.018

Hiemori, M., Bando, N., Ogawa, T., Shimada, H., Tsuji, H., Yamanishi, R., & Terao, J. (2000). Occurrence of IgE antibody-recognizing N-linked glycan moiety of a soybean allergen, Gly m Bd 28K. International Archives of Allergy and Immunology, 122(4), 238-245. https://doi.org/10.1159/000024404

Hirano, K., Hino, S., Oshima, K., Nadano, D., Urisu, A., Takaiwa, F., & Matsuda, T. (2016). Evaluation of allergenic potential for rice seed protein components utilizing a rice proteome database and an allergen database in combination with IgE-binding of recombinant proteins. Bioscience, Biotechnolology and Biochemistry, 80(3), 564-573. https://doi.org/10.1080/09168451.2015.1116927

Hu, W., Li, W., & Chen, J. (2017). Recent advances of microbial breeding via heavy-ion mutagenesis at IMP. Letters in Applied Microbiology, 65(4), 274-280.

International Atomic Energy Agency/Food and Agriculture Organization (IAEA/FAO) (n.d.). Mutant variety database. Retrieved from https://mvd.iaea.org

Ito, S. (2019). Contemporary global rice economies: Structural changes of rice production/consumption and trade. Journal of Nutritional Science and Vitaminology, 65, S23-S25. https://doi.org/10.3177/jnsv.65.s23

Ivanciuc, O., Schein, C.H., & Braun, W. (2003). SDAP: Database and computational tools for allergenic proteins. Nucleic Acids Research, 31(1), 359-362. https://doi.org/10.1093/nar/gkg010

Jensen-Jarolim, E., Schmid, B., Bernier, F., Berna, A., Kinaciyan, T., Focke, M., Ebner, C., Scheiner, O., & Boltz-Nitulescu, G. (2002). Allergologic exploration of germins and germin-like proteins, a new class of plant allergens. Allergy, 57(9), 805-810. ttps://doi.org/10.1034/j.1398-9995.2002.23686.x

Jespersen, M.C., Peters, B., Nielsen, M., & Marcatili, P. (2017). BepiPred-2.0: Improving sequence-based B-cell epitope prediction using conformational epitopes. Nucleic Acids Research, 45(W1), W24-W29. https://doi.org/10.1093/nar/gkx346

Katz, Y., Rajuan, N., Goldberg, M.R., Eisenberg, E., Heyman, E., Cohen, A., & Leshno, M. (2010). Early exposure to cow's milk protein is protective against IgE-mediated cow's milk protein allergy. Journal of Allergy and Clinical Immunology, 126(1), 77-82. https://doi.org/10.1016/j.jaci.2010.04.020

Kern, K., Havenith, H., Delaroque, N., Rautenberger, P., Lehmann, J., Fischer, M., Spiegel, H., Schillberg, S., Ehrentreich-Foerster, E., Aurich, S., & Treudler, R. (2019). The immunome of soy bean allergy: Comprehensive identification and

characterization of epitopes. Clinical and Experimental Allergy, 49(2), 239-251. https://doi.org/10.1111/cea.13285

Key, S., Ma, J.K., & Drake, P.M. (2008). Genetically modified plants and human health. Journal of the Royal Society of Medicine, 101(6), 290-298. https://doi.org/10.1258%2Fjrsm.2008.070372

Klausen, M.S., Jespersen, M.C., Nielsen, H., Jensen, K.K., Jurtz, V.I., Sønderby, C.K., Sommer, M.O.A., Winther, O., Nielsen, M., Petersen, B., & Marcatili, P. (2019). NetSurfP-2.0: Improved prediction of protein structural features by integrated deep learning. Proteins: Structure, Function, and Bioinformatics, 87(6), 520-527. https://doi.org/10.1002/prot.25674

Ko, D.K., Nadakuduti, S.S., Douches, D.S., & Buell, C.R. (2018). Transcriptome profiling of transgenic potato plants provides insights into variability caused by plant transformation. PloS One, 13(11), e0206055. https://doi.org/10.1371/journal.pone.0206055

Kringelum, J.V., Lundegaard, C., Lund, O., & Nielsen, M. (2012). Reliable B cell epitope predictions: Impacts of method development and improved benchmarking. PLoS Computational Biology, 8(12), e1002829. https://doi.org/10.1371/journal.pcbi.1002829

Koplin, J.J., Osborne, N.J., Wake, M., Martin, P.E., Gurrin, L.C., Robinson, M.N., Tey, D., Slaa, M., Thiele, L., Miles, L., & Anderson, D. (2010). Can early introduction of egg prevent egg allergy in infants? A population-based study. Journal of Allergy and Clinical Immunology, 126(4), 807-813. https://doi.org/10.1016/j.jaci.2010.07.028

Kuiper, H.A., Kleter, G.A., Noteborn, H.P., & Kok, E.J. (2001). Assessment of the food safety issues related to genetically modified foods. The Plant Journal, 27(6), 503-528. https://doi.org/10.1046/j.1365-313X.2001.01119.x

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Molecular Biology and Evolution, 35, 1547-1549. https://doi.org/10.1093/molbev/msy096

Lafarga, T., Wilm, M., Wynne, K., & Hayes, M. (2016). Bioactive hydrolysates from bovine blood globulins: Generation, characterisation and in silico prediction of toxicity and allergenicity. Journal of Functional Foods, 24, 142-155. https://doi.org/10.1016/j.jff.20 ZZ16.03.031

Laskowski, R.A., MacArthur, M.W., Moss, D.S., & Thornton, J.M. (1993). PROCHECK: A program to check the stereochemical quality of protein structures. Journal of Applied Crystallography, 26(2), 283-291. https://doi.org/10.1107/S0021889892009944

Latham, J.R., Wilson, A.K., & Steinbrecher, R.A. (2006). The mutational consequences of plant transformation. Journal of Biomedicine and Biotechnology, 25376. https://doi.org/10.1155/JBB/2006/25376

Li, Y., Liu, X., Zhu, Y., Zhou, X., Cao, C., Hu, X., Ma, H., Wen, H., Ma, X., & Ding, J.B. (2013). Bioinformatic prediction of epitopes in the Emy162 antigen of *Echinococcus multilocularis*. Experimental and Therapeutic Medicine, 6(2), 335-340.

Long, Y., Xu, W., Liu, C., Dong, M., Pei, X., Chen, R., Jin, W., Liu, W., & Li, L. (2020). Genetically modified soybean lines exhibit less transcriptomic variation compared to natural varieties. Research Square. https://doi.org/10.21203/rs.3.rs-112817/v1

López-Expósito, I., Song, Y., Järvinen, K.M., Srivastava, K., & Li, X.-M. (2009). Maternal peanut exposure during pregnancy and lactation reduces peanut allergy risk in offspring. Journal of Allergy and Clinical Immunology, 124(5), 1039-1046. https://doi.org/10.1016/j.jaci.2009.08.024

Matsaunyane, L.B., & Dubery, I.A. (2018). Molecular approaches to address intended and unintended effects and substantial equivalence of genetically modified crops transgenic crops-emerging trends and future perspectives. In M. S. Kwan & K. S. Malik (Eds.), Transgenic crops (pp. 115-131). United Kingdom: IntechOpen Limited.

Mermiri, D.Z.T., Lappa, T., & Papadopoulou, A.L. (2017). Review suggests that the immunoregulatory and anti-inflammatory properties of allergenic foods can provoke oral tolerance if introduced early to infants' diets. Acta Paediatrica, 106(5), 721-726. https://doi.org/10.1111/apa.13778

Montero, M., Coll, A., Nadal, A., Messeguer, J., & Pla, M. (2011). Only half the transcriptomic differences between resistant genetically modified and conventional rice are associated with the transgene. Plant Biotechnology Journal, 9(6), 693-702. https://doi.org/10.1111/j.1467-7652.2010.00572.x

National Center for Biotechnology Information (NCBI). (1988). GenBank. Bethesda, United States: National Library of Medicine (US), National Center for Biotechnology Information.

Opaleye, O., Rose, R.S., Whittaker, M.M., Woo, E.J., Whittaker, J.W., & Pickersgill, R.W. (2006). Structural and spectroscopic studies shed light on the mechanism of oxalate oxidase. Journal of Biological Chemistry, 281(10), 6428-6433. https://doi.org /10.1074/jbc.m510256200

Ricroch, A.E. (2013). Assessment of GE food safety using '-omics' techniques and long-term animal feeding studies. New Biotechnology, 30(4), 349-354. https://doi. org/10.1016/j.nbt.2012.12.001

Ricroch, A.E., Bergé, J.B., & Kuntz, M. (2011). Evaluation of genetically engineered crops using transcriptomic, proteomic, and metabolomic profiling techniques. Plant Physiology, 155(4), 1752-1761. https://doi.org/10.1104/pp.111.173609

Santino, A., Taurino, M., De Domenico, S., Bonsegna, S., Poltronieri, P., Pastor, V., & Flors, V. (2013). Jasmonate signaling in plant development and defense response to multiple (a)biotic stresses. Plant Cell Reports, 32(7), 1085-1098. https://doi.org/10.100 7/s00299-013-1441-2

Schnell, J., Steele, M., Bean, J., Neuspiel, M., Girard, C., Dormann, N., Pearson, C., Savoie, A., Bourbonniere, L., & Macdonald, P. (2015). A comparative analysis of insertional effects in genetically engineered plants: Considerations for pre-market assessments. Transgenic Research, 24(1), 1-17. https://doi.org/10.1007/s11248-014-9843-7

Sikic, K., Tomic, S., & Carugo, O. (2010). Systematic comparison of crystal and NMR protein structures deposited in the protein data bank. The Open Biochemistry Journal, 3(4), 83-95. https://doi.org/10.2174/1874091x01004010083

Sormanni, P., Aprile, F.A., & Vendruscolo, M. (2015). Rational design of antibodies targeting specific epitopes within intrinsically disordered proteins. Proceedings of the National Academy of Sciences USA, 112(32), 9902-9907. https://doi.org/10.1073/pnas.1422401112

Tahmoorespur, M., Nazifi, N., & Pirkhezranian, Z. (2017). In silico prediction of B-cell and T-cell epitopes of protective antigen of *Bacillus anthracis* in development of vaccines against anthrax. Iranian Journal of Applied Animal Science, 7(3), 429-436.

Ververis, E., Ackerl, R., Azzollini, D., Colombo, P.A., de Sesmaisons, A., Dumas, C., Fernandez-Dumont, A., da Costa, L.F., Germini, A., Goumperis, T., & Kouloura, E. (2020). Novel foods in the European Union: Scientific requirements and challenges of the risk assessment process by the European Food Safety Authority. Food Research International, 137, 109515. https://doi.org/10.1016/j.foodres.2020.109515

Wang, P., Sidney, J., Dow, C., Mothé, B., Sette, A., & Peters, B. (2008). A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. PLoS Computational Biology, 4(4), e1000048. https://doi.org/10.1371/journal.pcbi.1000048

Wang, P., Sidney, J., Kim, Y., Sette, A., Lund, O., Nielsen, M., & Peters, B. (2010). Peptide binding predictions for HLA DR, DP and DQ molecules. BMC Bioinformatics, 11(1), 568. https://doi.org/10.1186/1471-2105-11-568

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R., & Schwede, T. (2018). SWISS-MODEL: Homology modelling of protein structures and complexes. Nucleic Acids Research, 46(W1), W296-W303. https://doi.org/10.1093/nar/gky427

Weinhold, A. (2018). Transgenerational stress-adaption: An opportunity for ecological epigenetics. Plant Cell Reports, 37(1), 3-9. https://doi.org/10.1007/s00299-017-2216-y

Yuan, S., Chan, H.S., Filipek, S., & Vogel, H. (2016). PyMOL and Inkscape bridge the data and the data visualization. Structure, 24(12), 2041-2042. https://doi.org/10.1016/j.str.2016.11.012

Zimmermann, G., Bäumlein, H., Mock, H.-P., Himmelbach, A., & Schweizer, P. (2006). The multigene family encoding germin-like proteins of barley. Regulation and function in basal host resistance. Plant Physiology, 142(1), 181-192. https://doi.org/10.1104/pp.106.083824