

Identification and annotation of peptide allergens in *Prunus dulcis*

Arshwinder Singh

Thapar University

Atul Kumar Upadhyay (✉ atul.upadhyay@thapar.edu)

Thapar University

Research Article

Keywords: Food allergy, Allergens, Protein, Genome, Prediction, Functional annotation

Posted Date: April 5th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1507943/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Approximately, 10% of the world population is facing the challenge of food allergy in direct or indirect way. In this study, a genome-wide identification and annotation of the novel putative allergen from Almond is performed. Initially, the whole proteome of Almond (31,000 proteins) was scanned by Allergenonline, a publically available database of already reported allergens from different sources. The detailed analysis suggests that there are 430 putative allergens which reduced to 45 on motif-based screening using AllFam database. These predicted allergens are annotated for their function by using PFAM, GO databases and orthology analysis. To validate our prediction, we have used structural insights of allergen and antibody interactions for one of the predicted putative allergen protein, homologous to Pru ar 3.0101 allergen from Apricot. The structure of putative allergen was modeled and molecular docking studies were performed against the antibody. The best docked conformation was subjected to molecular simulation studies to confirm the stable binding of these two molecules. This detailed analysis suggests that the identified allergen will show cross reactivity similar to Pru ar 3.0101 allergen from Apricot. This is one of the first report of identifying and annotating the homologous of Pru ar 3.0101 allergen in Almond.

Introduction

Prevalence of food allergy ranges from five to ten percent in the world population (Pepper, 2020) with no current cure (Moore, Stewart, & deShazo, 2017). Tree nuts are allergic foods that often cause deadly anaphylactic reactions. Generally, the tree nut allergies that develops in childhood continues to adulthood (Iweala, 2018). As estimated, tree nut allergies contribute by affecting about 0.05 to 7.3 % of the population (Weinberger T, 2018). Tree nut allergens mostly constitute seed storage proteins such as vicilins and legumins. While in an allergic reaction, antibody IgE present on the surface of mast cells and basophils bind to tree nut allergens, resulting in degranulation and release of histamine and other inflammatory mediators (Smeekens JM, 2018). Tree nuts, also contain so-called pan-allergens (profilins, heveins, lipid transfer proteins) which show high IgE-mediated cross-reactivity towards pollen and food homologs (Valenta R, 1992). Reports from PREVALE (Prevalence of Food Allergy in Leganés) reveals that the cases of allergy to tree nuts in the first 3 years of lifespan, alone accounts to 0.83% of the total (LÁ, 2020).

Till now individual tree nut allergies have receive only minimal public attention as compared with peanut allergies though pervasiveness of these is equally high as that of peanut allergy, as a reason, tree nut allergies are not well studied (Che H, 2019) To exemplify, there are only 2 current macadamia (*Macadamia integrifolia*) nut allergens identified though several such allergies have been recorded.

Besides this, tree nuts have occupied major portion of current human diets mainly due to their great taste and promising health benefits. Various studies have reported the beneficial impacts of tree nuts on coronary diseases and cholesterol serum levels (Rehm & Drewnowski, 2017). Among all the tree nuts, almonds (*Prunus dulcis* or *Amygdalus communis* L.) constitute a prime portion of global consumption,

followed by walnuts (*Juglans regia*), cashews (*Anacardium occidentale*), pistachio nuts (*Pistacia vera*), hazelnuts (*Corylus avellana*), pecan nuts (*Carya illinoensis*), macadamia nuts (*Macadamia ternifolia*), brazil nuts (*Bertholletia excelsa*), and pine nuts (*Pinus pinea* and other *Pinus* species). As a consequence, being highly consumed there are various reported cases of almond related allergies.

Reports from United States, stated the prevalence of almond and cashew nut allergies to be 0.7% among children and adults (Gupta RS, 2018). Likewise, Australian survey affirmed the prevalence of almond allergy to be 0.2% in adolescents (Sasaki M, 2018) and 0.3% in children (McWilliam V, 2019). Peanut allergy victims demonstrated a high rate of concurrent almond sensitization ranging from 45 to 71%, however only 11% showed allergenic symptoms (Mustafa SS, 2020). Patients with tree nut allergies showed almond sensitization of about more than 40% (Elizur A, 2018).

Currently, there are six almond (*Prunus dulcis*) allergens that have been inscribed in the WHO/IUIS database of allergens, namely Pru du 3: a non-specific lipid transfer protein 1, Pru du 4: a profilin, Pru du 5: a 60s acidic ribosomal prot, Pru du 6: Amandin, 11S globulin legumin-like protein, Pru du 8: antimicrobial seed storage protein, Pru du 10: mandelonitrile lyase 2.

The foremost step for recognizing the allergy related impacts of food proteins is, discovering and characterizing food allergens (Che H, 2019). The aim of present study is to screen novel allergen proteins by the proteome wide analysis of almond species (*Prunus Dulcis*). We have filtered out the putative allergens on the basis of sequence identity to already known allergens. Thereafter, motif analysis along with homology studies further refined our set of putative allergens. Finally, T-cell and B-cell epitope conservancy examination led us to a very narrow set of putative allergens.

Materials And Methods

Protein sequence resources

The Almond proteome (*Prunus dulcis*, Strain: cv. Texas) was administered from publicly available database namely, UniPort (www.uniprot.org). It constitutes 8 chromosomes with 6618, 3668, 3486, 3294, 2745, 4196, 3166, 3074 number of proteins present in respective chromosomes and 1764 protein as unassembled. This totals to 31,932 proteins in the complete proteome of almond (Alioto, 2020) .

Identification of potential allergen proteins

For discovering potential allergens from *Prunus Dulcis* proteome, Allergenonline, version 21 was used (www.allergenonline.org). This database contains 2233 protein sequences (913 groups, 430 species) and is maintained by Food Allergy Research and Resource Program (FARRP), University of Nebraska–Lincoln. For identification the FASTA sequence was searched against the database with search method Full Fasta 36, provided within the Allergenonline toolbox. Credentials: E value $\leq 1E-07$ and Identity $\geq 50\%$ were used while searching and max alignments was set to default.

Motif based screening of potential cross-reactive protein sequences

Identified potential allergens were further screened. In, here we selected allergen sequences from AllFam database and used them as a reference for screening. AllFam constitute 151 families with 1059 allergens in total and is maintained by Medical University of Vienna. Reference allergens were chosen on two bases, first on the basis of their routes of exposure, second was the source. Motifs from query sequence were compared to that found in reference and the sequences that matched the most were selected

Classification of putative allergens into protein families along with their gene ontology (GO)

The retrieved sequences were subjected to Pfam database (www.pfam.xfam.org) and Conserved Domain Database (www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) for protein family analysis. Pfam 34.0 contains 19179 families (as of May 2021) and was produced and maintained by European Bioinformatics Institute. Protein families were retrieved by using Hidden Markov Models Method (www.ebi.ac.uk/Tools/hmmer/search/hmmscan). Gene ontology (GO) was carried out to determine the biologic, molecular or cellular function of all the discovered potential allergens. GO accession numbers of retrieved allergens were obtained from InterPro database (www.ebi.ac.uk/interpro). For plotting and visualizing the ontology data the GO accessions were submitted to Web Gene Ontology Annotation Plot (WEGO, www.wego.genomics.cn).

Orthology Modelling and Phylogeny

Galaxy server was used for analyzing orthologs of putative allergens (<https://usegalaxy.eu/>).

Proteinortho tool from galaxy server was used for finding the orthologs (Lechner, 2011). It is a tool used to detect orthologous genes or proteins within different species. Basically, it equates the similarities between the given data sequences (genes or proteins) and assemble them to significant groups. Diamond Algorithm and E-value $1E-03$ was used for finding the orthologs. Orthology analysis was carried out using the proteomes of related species i.e., Apricot (*Prunus armeniaca*), Cherry (*Prunus avium*) and Peach (*Prunus persica*). Orthology groups with allergenic proteins were examined and selected. Phylogeny tree of the selected sequences was built using Phylogeny.fr (Dereeper A., 2010).

B-cell epitope prediction

The Immune Epitope Database (IEDB) was used for prediction of B cell epitopes (<https://www.iedb.org/>). IEDB constitute experimental data of antibody and T cell epitope studies conducted in humans and other animals. Majority of the available data is from infectious disease, allergy and autoimmunity studies. B-cell linear epitopes were predicted using Bepipred 2.0 and ABCpred (Nemati Zargar, 2021). Bepipred 2.0 web server uses Hidden Markov Model (HMM) whereas ABCpred uses artificial neural network (ANN) to predict B cell epitopes. Thereafter, to find the conserved epitopes among the allergens and non/putative-allergen proteins, epitope conservancy analysis was carried out.

Structure prediction by molecular modelling

Swiss model server was used for structure modelling of the selected potential allergens (www.swissmodel.expasy.org/). Swiss model is a web server maintained by Swiss Institute of

Bioinformatics Biozentrum at University of Basel. The target sequence was uploaded along with project title to search for templates. Among all the suggested templates a specific template on the basis of its coverage, GMQE (Global Model Quality Estimate), identity and method of structure prediction was selected to predict the model of target sequence. The Global quality estimate, local quality estimate and Z-score index of the predicted model was used to assess the quality of predicted model. Quality of structure was also assessed by ProCheck tool. Model with best quality index was selected for further analysis.

Molecular Docking

Cluspro 2.0, server was used for molecular docking (www.cluspro.bu.edu/home.php). It is an online server created and maintained by Vajda lab at Boston University. Finally selected potential allergen structure was uploaded as a ligand whereas an antibody structure procured from protein data bank was uploaded as receptor onto the server. For specific antibody-protein docking, antibody mode provided by Cluspro was used. For precise results, non-CDR regions of antibody were masked on.

Molecular dynamic simulation studies

MD simulations of the best docked results with most negative energy was carried out using GROMACS 2020.1 software package. The elements of the simulation system were protein and water. All the MD simulations were executed using CHARMM36 force field (Huang, 2016). A triclinic box was created around the protein-protein complex and thereafter solvated with water model TIP3P. The system was positively charged and to neutralize, 3 CL negative ions were added. After Energy minimization, position restrained was applied followed by NVT and NPT equilibrium. NVT (constant Number of particles, Volume, and Temperature) was performed with coupling groups of protein and non-protein, at temperature of 300 K, with a coupling constant of 0.1 ps for 100 ps whereas for NPT (Number of particles, Pressure, and Temperature are all constant) pressure of 1 bar with coupling constant of 2.0ps for 100 ps with the same coupling groups was applied. Finally, MD simulation of 25 ns was executed (Lemkul, 2018).

Results

Identification of putative allergens

Full length sequence alignment was performed to identify putative allergens. Newly submitted proteome was procured from UniProt. It constitutes 31,934 proteins, with 31932 as unreviewed and 2 as reviewed. We compared our query sequences to allergenonline database and found 430 sequences as putative allergens. The allergenonline toolbox uses scoring matrix of BLOSUM 50 to execute these alignments. As BLOSUM 50 is believed to screen highly similar proteins in terms of their overall structure and function, irrespective of their evolutionary relation.

Screening of potential allergenic sequences in Prunus Dulcis

Identified allergenic sequences were further screened by comparing with reference from AllFam Database. Database has categorized allergens into various families with their sources and routes of exposure. A reference was created from the proteins in AllFam database on the basis of their allergy causative method i.e., ingestion, contact and inhalation.

Motifs plays an important role in allergenicity of an allergen. There were 229 allergen sequences selected for reference. Motifs for query allergen sequences as well as reference allergen sequences were identified and compared using motif finder from GenomeNet server (www.genome.jp/tools/motif/). Finally, the query sequences with matched motifs were selected and this amounted to 45. These sequences were further characterized and analyzed for allergenicity.

Protein Family And Gene Ontology Analysis

Screened query sequences were subjected to protein family and gene ontology analysis. Conserved Domain Database and Pfam database were used for protein family analysis. Out of 45 query proteins maximum number of proteins were found in GST_N (Glutathione S-transferase) and this amount to 10. This family includes phase II enzymes with capability to perform conjugation reactions. Bet_v_1 contained the second most number of proteins with 5 in number followed by EF-hand_1 with 4 proteins (Fig. 1A). EF-hand motifs play their role in calcium binding properties of proteins. On the other hand, protein families NmrA, Enolase_C, Thioredoxin GST_N_3 and Thaumatin all constitute equal number of proteins i.e., three out of 45 putative allergens 35 has GO annotations and these constitute to 66 different GO terms which were categorized into cellular component, molecular function and biological process. In case of cellular component 8.6% of 45 proteins were located in cell part (GO:0044464), 8.6% in cell (GO:0005623) and another 8.6% of genes were located in protein-containing complex (GO:0032991). Considering molecular function 57.1% are seen to show binding activity (GO:0005488) whereas 28.6% own catalytic activity (GO:0003824). While in case of biological process 65.7% of protein are involved in metabolic processes (GO:0008152), 65.7% depicted their role cellular process (GO:0009987) and only a little i.e., 5.7% of them take part in localization (GO:0051179) (Fig. 1B).

Orthology and phylogeny analysis

Protein family and gene ontology characterized protein sequences were subjected to ortholog analysis. All the 45 protein sequences were put through ortholog analysis by means of a tool i.e., Proteinortho. Total 45 ortholog groups were created. Among all, 5 groups included orthologs that are already reported allergens and are listed on WHO/IUIS Allergen Nomenclature. Phylogeny tree depicts the groups with their respective orthologs. Accession numbers highlighted in red color belong to the category of already reported allergens and the one with blue color belong to the category of putative allergens. There were total of 4 motifs found namely, Bet_v_1 (Pathogenesis-related protein Bet v 1 family), Thaumatin, Profilin and LTP_2 (lipid transfer protein). Highlighted (blue) proteins showed more than 90% similarity to their respective ortholog allergen protein, hence were considered for further analysis (Fig. 2).

B-cell epitope prediction

There were 7 putative allergens considered for B-cell epitope analysis. The given table illustrates the predicted epitopes with their positions. Default parameters were used while predicting the linear B-cell epitopes. Primarily, epitopes of all the allergens were predicted along with the epitopes of their respective ortholog allergens. Epitopes that were 100% conserved in both i.e., orthologous reported allergens and our almond putative allergens, were screened and used for further analysis (Table 1).

Table 1
B-cell epitopes of screened putative allergens

Putative allergens	Predicted epitopes	Position	Conserved epitopes
A0A5E4EF81	26–29, 42–51, 62–69, 81–93, 100–129	114–130	AKYLDMGQVFSAADNI
A0A5E4E9E4	8–19, 54–69, 84–88	13–20,	MCEIEGN
A0A5E4F9Q4	39–57, 90–96, 125–132	42–57	AEIVEGDGGVGTIKKI
A0A5E4FC44	41–57, 90–96, 124–132	42–57	AEIVEGDGGVGTIKKI
A0A5E4FGB4	41–57, 73–79, 125–132	42–57	AEIVEGDGGVGTIKKI
A0A5E4GAL7	47–58, 84–92, 94–113, 149–177, 187–214, 217–232	47–58, 84–92, 94–113, 159–177, 187–212, 217–232	DQKPQLSLTGFE, STDASGKFT, ATADCGSGQVSCNGNGAAPP, GTGECKASTCPADINKVCPAPLQVKGSDG, LAFNQPKYCCTPPNDKPETCPPTDYSKL, TQCPQAYSAYDDKSS
A0A5E4GM68	44–61, 95–111, 162–173, 183–205, 217–228	44–55, 95–111, 183–205	DQKPQLSTTGFE, CASGQVTCNGAIPPA, CVKFGEPRYCCTPPQNTPETCPP

Structure Modelling of selected protein sequences

Among 45 putative allergen sequences 7 proteins that showed more than 90% identity to their allergen ortholog were considered structure modelling. These proteins were modelled using Swiss model server. Templates were searched against the PDB (Protein Data Bank). 2b5s.1 (Pru p3, the prototypic member of the family of plant non-specific lipid transfer protein) template was used for modelling of putative allergen protein (A0A5E4EF81). This showed coverage of 0.68%, sequence identity of 92.31% and GMQE of 0.91. Structure quality parameters with 0.48 GMQE scores, QMEAN of 0.77, Z-score lied close to 0 and Local quality estimate scores more than 0.6 confirmed the quality of the modelled structure.

PROCHECK analysis provided the Ramachandran plot which also confirmed the quality of modelled structures. It depicted about 97.2% of residues in the structure lies in the most favored regions 2.8% in additional allowed regions and no residues were found in disallowed regions. Moreover, in case of A0A5E4E9E4, template used is 5fds (monomeric allergen profilin (Hev b 8)) and Ramachandran plot shows the presence of 91.5% of residues in the structure lies in the most favored regions 5.6% in additional allowed regions and 2.8% in generously allowed regions. For A0A5E4FGB4, A0A5E4FC44, A0A5E4F9Q4 proteins, 6stb.1.A (strawberry pathogenesis-related 10 (PR-10), Fra a 1.02 protein, Q64W mutant) template is used. Ramachandran plot all the 3 proteins illustrates presence of 91.3% of residues in most favored regions and 8.7% in additional allowed regions. 3zs3.1.A (Mal d 2, the thaumatin like food allergen from apple) template is used for modelling the structure of A0A5E4GAL7 and 2ahn.1.A (cherry allergen Pru av 2) is used for A0A5E4GM68 protein. Ramcahndran plot of structure of A0A5E4GAL7 and A0A5E4GM68 demosntrate the presence of 88.9% and 91.1% residues in most favored regions, 11.1% and 8.9% residues in additional allowed regions respectively (Fig. 3).

Molecular Docking And Molecular Dynamic Simulation Studies

Antibody structure procured from PDB was docked against the modelled structure. Antibody mode of Online server Cluspro 2.0 was used for docking. Docking generated various docked models and the best docked model with lowest energy was selected for further analysis (Table 2). Rainbow color sites in the figure depicts the predicted linear conserved B-cell epitopes (Fig. 4A). Lowest binding energy of all the docking models

Table 2
Lowest binding energy of all the docking models

Number of Models	Binding Energy (Kcal/mol)
A0A5E4EF81	-362.2
A0A5E4F9Q4	-338.4
A0A5E4FC44	-332.3
A0A5E4GM68	-332.2
A0A5E4GAL7	-327.3
A0A5E4FGB4	-323.4
A0A5E4E9E4	-290.6

The conformational stability of the antigen-antibody complex during MD simulation was analyzed by root mean square deviation (RMSD). RMSD was calculated for backbone atoms. At most points throughout simulation, the system laid within the RMSD window of 0.05-0.9 nm. At the beginning, RMSD

rose un-evenly until 0.9 ns, thereafter it became begun to decline to reach at 0.4 nm at the end of the end of simulation (Figure 4B).

Discussion

Almonds are one of the most consumable tree nuts among all, yet very little is known about its cross-reactive proteins. Hence, screening of allergenic proteins from almonds is obligatory to provide new insights to the rising almond related allergies. In this study we have attempted a genome-wide identification approach to find putative almond allergens. The process of gauging a protein to be allergenic was formulated by the Codex Alimentarius Guidelines in 2003 and 2009. It suggested a pre-identification step as the comparison of query protein sequences with already existing allergens by means of informatics tools. We used allergenonline database to blast our query proteins to already known allergens and found about 430 putative allergen sequences. For further screening we created a reference out of allergen proteins present in AllFam database. Though AllFam database has listed allergens on the basis of various routes of exposure such as auto-allergens, contact, iatrogenic, ingestion, inhalation, parasite infestation, sting, bite (Radauer, 2017). But as our query allergen is an eatable tree nut, so the allergens that leads to allergy by ingestion, contact and inhalation, were selected as reference from the database. This was necessary for further screening as allergenonline database contains all types of allergens with varied modes of exposure. Motifs were identified from both, putative allergens and reference and after comparing the two we selected the sequences with identical motifs. This gave us 45 proteins out of 430. We further classified the sequences to their respective families and gene ontology using pfam and InterPro databases respectively. Furthering, the putative allergen sequences were subjected to orthology analysis. The closely related species including Apricot (*Prunus armeniaca*), Cherry (*Prunus avium*) and Peach (*Prunus persica*) were selected for orthology analysis. This provided us with 7 putative allergens from 45. The basis for selecting these 7 proteins, was their more than 90% sequence identity along with identical motifs in same orthologous allergenic proteins. We used SWISS-MODEL server to select templates and model the structures of selected proteins, as their structure has not been resolved by anyone earlier and so was not present at protein data bank (Waterhouse, 2018). Later, the modelled structures were assessed for their quality by SWISS-MODEL scores and additionally by PROCHECK tool. B-cell epitopes for the modelled structure were predicted and the structures were subjected to docking. Finally, the best docked complex.i.e. A0A5E4EF81 (lipid transfer protein) was selected and subjected to molecular dynamics studies (MDS). Simulation was carried out for 25 ns, by means of GROMACS.

Lipid transfer proteins (LTP) are antigens responsible for allergenic reactions to variety of entities including pollen, latex and certain plant foods. These proteins are stabilized by disulphide bonds enabling them with a resistance to digestion, hence letting them to reach intestine to cause sensitization and allergic reactions (Tordesillas L, 2013). Pru 3 (peach allergen) a lipid transfer protein (LTP) is a major allergen leading to severe allergic reactions in Mediterranean people (Flores E, 2012). Jug r 3 (walnut allergen) is another allergen belonging to LTP family, known to effect Italian population (Pastorello, 2004).

Allergenicity to a particular food can also lead to cross-reaction to a closely related food. Generally, cross-reactivity among tree nut allergens is very high especially between cashew and pistachio being from the same family, i.e. Anacardiaceae family, walnut and pecan from Juglandaceae family (Andorf S, 2017). Jug r 5, Cor a 1, Ara h 8 from walnut, hazelnut and peanut respectively, have been recognised to show cross-reaction to Bet v1 (Wangorsch A, 2017) (Mittag D, 2004) (Hofmann C, 2013). Similarly, Pru p 3 (peach lipid transfer protein) has been identified to show cross reactivity to certain lipid transfer proteins including Ara h 9 from peanut, Cor a 8 from hazelnut, Jug r 3 from walnut and Pru du 3 from almond (Mothes-Luksch N, 2017).

Identity of protein sequences along with structural epitopes play an important role in predicting cross-reactivity (Yu, 2022). For protein sequence identity, it has been observed that there is at least 70% sequence identity between two cross-reacting allergens (RC, 2000). For more precise prediction of cross-reactivity, IgE binding epitopes along with the location of identical sequences should be considered (Smeekens JM, 2018). The usage of amino acids in the translation of proteins and peptide of different allergen families also have unique patterns (Singh, 2022).

Considering the suggested possibilities, we predicted a putative cross-reactive allergen, i.e., A0A5E4EF81 a lipid transfer protein. This peptide shows 92.11% sequence identity to Pru ar 3.0101 an allergen from *Prunus armeniaca* (Apricot) along with conserved B-cell epitopes. Docking and MDS analysis (stable RMSD) proves the stability of antigen-antibody complex, hence this confirms the possibility of our predicted peptide to be cross-reactive to Pru ar 3.0101, however, there is still a need for *in vitro* verification.

Conclusion

In the present study, we executed genome-wide identification of a novel putative allergen from almonds. Screening of protein sequences from proteome through AllergenOnline and AllFam database along with ortholog analysis has narrowed us to a single protein which was further subjected to docking and molecular dynamics studies analysis to confirm its cross-reactivity. Main aim of this study, was to provide new insights into almond-associated allergy. As, almonds are among the most consumable tree nuts throughout the world, so there is a special need to safeguard the consumption of these nuts.

References

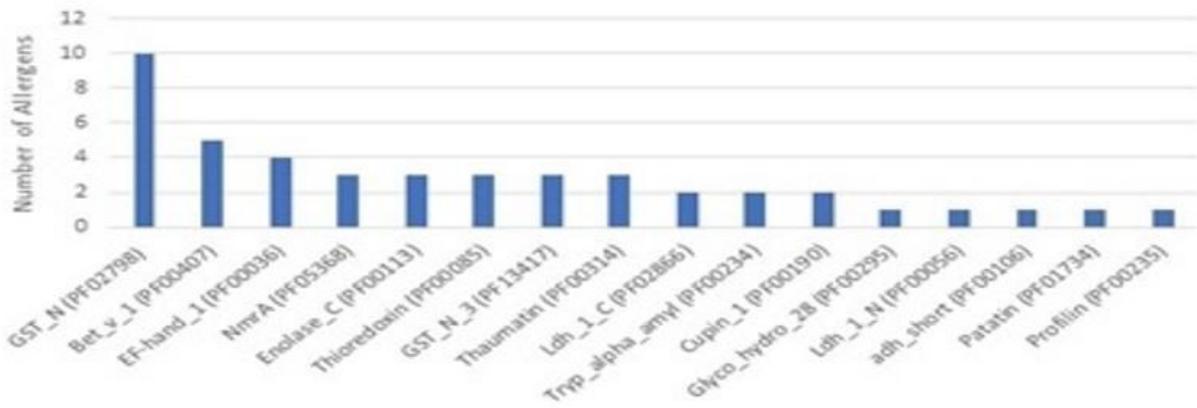
1. Alioto TA-G (2020) Transposons played a major role in the diversification between the closely related almond and peach genomes: results from the almond genome sequence. *Plant J* 101:455–472
2. Andorf S, B. M (2017) Association of Clinical Reactivity with Sensitization to Allergen Components in Multifood-Allergic Children. *J Allergy Clin Immunol Pract*, 1325–34
3. Che H, Z. Y (2019) Identification of Almond (*Prunus dulcis*) Vicilin As a Food Allergen. *J Agric Food Chem*, 425–432

4. Dereeper A, A. S (2010) BLAST-EXPLORER helps you building datasets for phylogenetic analysis. *BMC Evol Biol*
5. Elizur A, A. M (2018) NUT co reactivity - ACquiring knowledge for elimination recommendations (NUT CRACKER) study. *Allergy Eur J Allergy Clin Immunol* 73:593–601
6. Flores E, C. L.-P (2012) Plant food allergy in patients with pollinosis from the Mediterranean area. *Int Arch Allergy Immunol*,346–54
7. Gupta RS, W. C (2018) The public health impact of parent-reported childhood food allergies in the United States. *Pediatrics* 142:e20181235
8. Hofmann C, S. S (2013) Cor a 1-reactive Tcells and IgE are predominantly cross-reactive to Bet v 1 in patients with birch pollen-associated food allergy to hazelnut. *J Allergy Clin Immunol*,1384–92
9. Huang JR (2016) CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat Methods* 14:71–73
10. Iweala OC (2018) Food Allergy. *Curr Gastroenterol Rep* 20:17
11. LÁ E-Z (2020) TREE NUTS ALLERGY: Knowledge, gaps and practical implications. *Allergol Immunopathol (Madr)*,313–315
12. Lechner MF (2011) Proteinortho: Detection of (Co-)orthologs in large-scale analysis. *BMC Bioinformatics*,124
13. Lemkul J (2018) From Proteins to Perturbed Hamiltonians: A Suite of Tutorials for the GROMACS-2018 Molecular Simulation Package, v1.0. *Living J. Comp. Mol. Sci*
14. McWilliam V, P. R (2019) Patterns of tree nut sensitization and allergy in the first 6 years of life in a population-based cohort. *J Allergy Clin Immunol* 143:644–650
15. Mittag D, A. J.-W (2004) Ara h 8, a Bet v 1-homologous allergen from peanut, is a major allergen in patients with combined birch pollen and peanut allergy. *J Allergy Clin Immunol*,1410–7
16. Moore LE, Stewart PH, deShazo RD (2017) Food Allergy: What We Know Now. *Am J Med Sci* 353:353–366
17. Mothes-Luksch N, R. M.-T.-F (2017) Pru p 3, a marker allergen for lipid transfer protein sensitization also in Central Europe. *Allergy*,1415–8
18. Mustafa SS, V. K (2020) Real-world tree nut consumption in peanut-allergic individuals. *Ann Allergy Asthma Immunol* 124:277–282
19. Nemati Zargaran FA (2021) Detecting the Dominant T and B Epitopes of Klebsiella pneumoniae Ferric Enterobactin Protein (FepA) and Introducing a Single Epitopic Peptide as Vaccine Candidate. *Int J Pept Res Ther*,2209–2221
20. Pastorello EA (2004) Lipid transfer protein and vicilin are important walnut allergens in patients not allergic to pollen. *Journal of Allergy and Clinical Immunology*,908–914
21. Pepper AN (2020) Consensus report from the Food Allergy Research & Education (FARE) 2019 Oral Immunotherapy for Food Allergy Summit. *J Allergy Clin Immunol* 146(2):244–249

22. Radauer C (2017) Navigating through the Jungle of Allergens: Features and Applications of Allergen Databases. *Int Arch Allergy Immunol* 173:1–11
23. RC A (2000) Structural biology of allergens. *J Allergy Clin Immunol.*, 228 – 38
24. Rehm CD, Drewnowski A (2017) Replacing American snacks with tree nuts increases consumption of key nutrients among US children and adults: Results of an NHANES modeling study. *Nutr. J*,16,17
25. Sasaki M, K. J (2018) Prevalence of clinic-defined food allergy in early adolescence: the SchoolNuts study. *J Allergy Clin Immunol* 141:391–398
26. Singh BA (2022) Structural insights into the amino acid usage variations in the profilin gene family. *Amino Acids*,411–419
27. Smeekens JM, B. K (2018) Tree nut allergies: Allergen homology, cross-reactivity, and implications for therapy. *Clin Exp Allergy*,762–772
28. Tordesillas L, G.-C. C.-A.-G (2013) Transport of Pru p 3 across gastrointestinal epithelium an essential step towards the induction of food allergy? *Clin Exp Allergy*,1374–83
29. Valenta R (1992) D. M. Profilins constitute a novel family of functional plant pan-allergens. *J Exp Med*, 377 – 85
30. Wangorsch A, J. A.-W (2017) Identification and implication of an allergenic PR-10 protein from walnut in birch pollen associated walnut allergy. *Mol Nutr Food Res.*
31. Waterhouse AB (2018) SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res*,W296-W303
32. Weinberger T, S. S (2018) Current perspectives on tree nut allergy: a review. *J Asthma Allergy* 11:41–51
33. Yu PZ (2022) Immunoinformatics Construction of B Cell Epitope-Based Hypoallergenic Der f 34 Vaccine for Immunotherapy of House Dust Mite Allergy. *Int J Pept Res Ther*

Figures

1A.



1B.

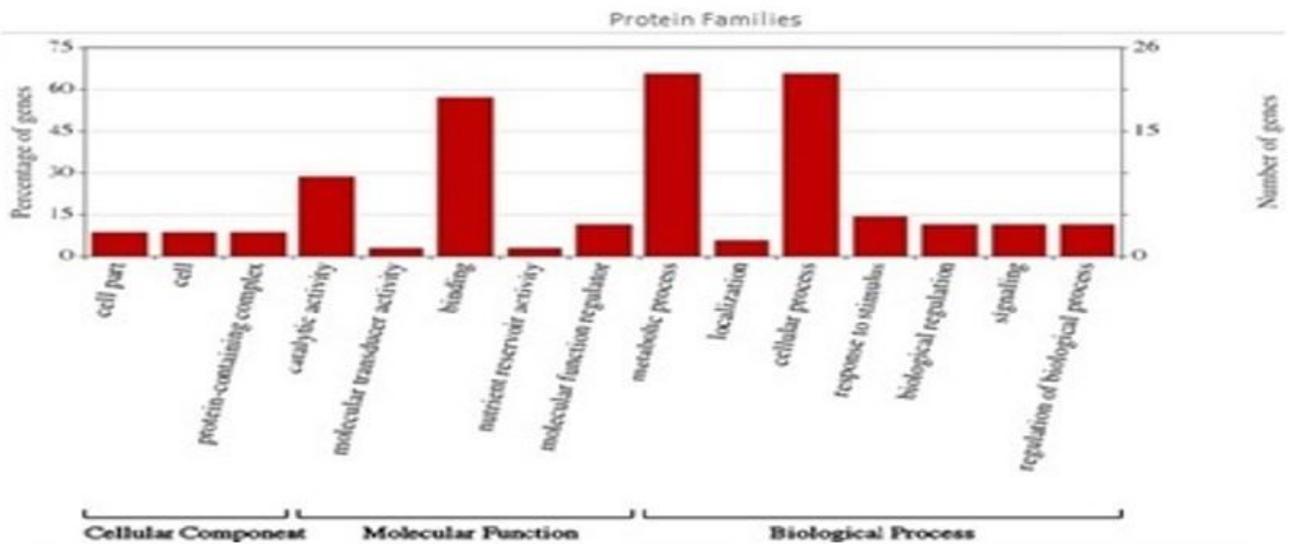


Figure 1

A). Classification of putative allergens into different protein families. The 16 most prominent protein families based on number of allergens present in the particular protein family are plotted. B). Gene Ontology (GO term) analysis of putative allergen proteins and the results are plotted on the basis of percentage and number of allergens present in different GO terms.

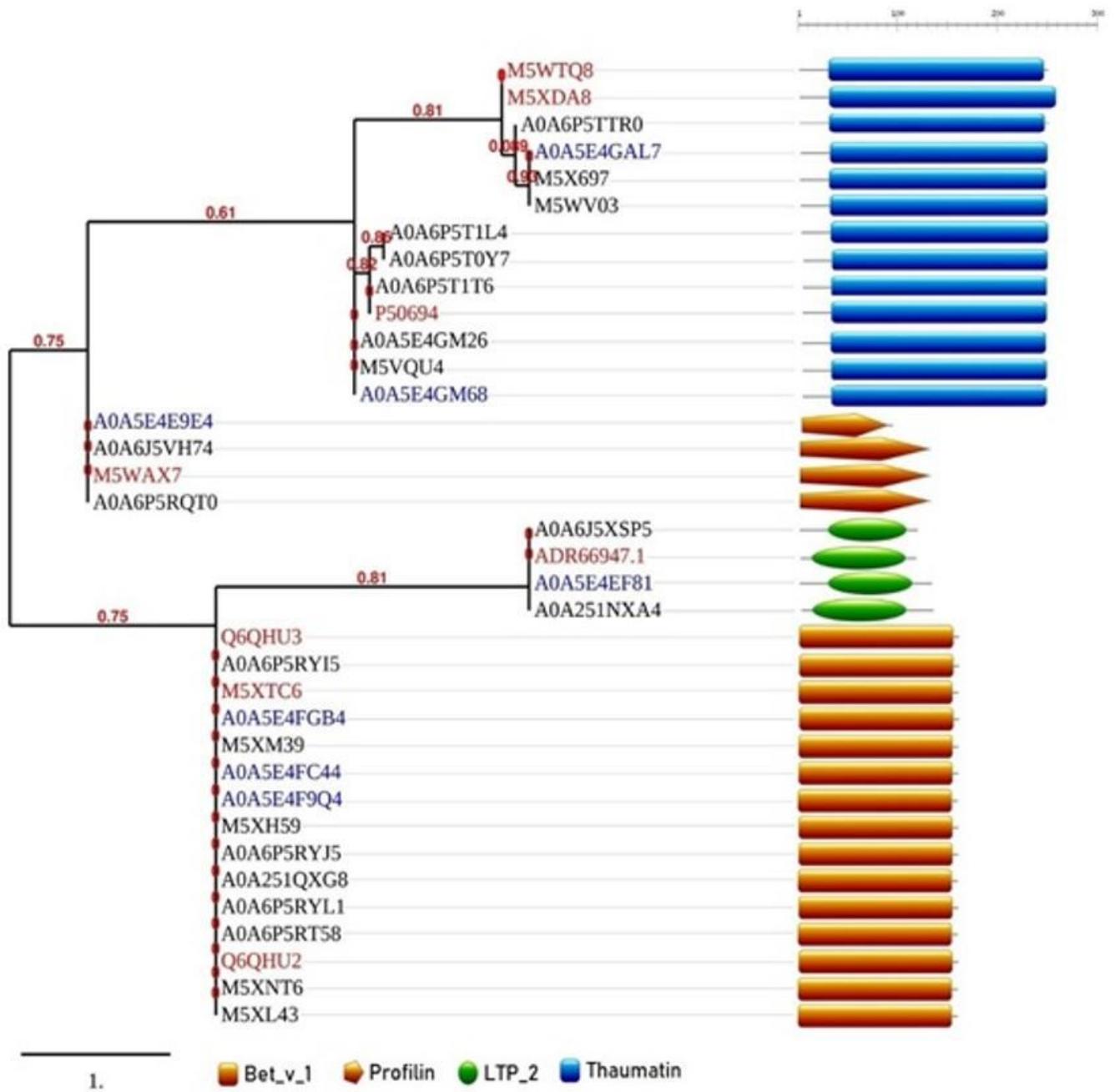


Figure 2

Orthology analysis of putative allergens

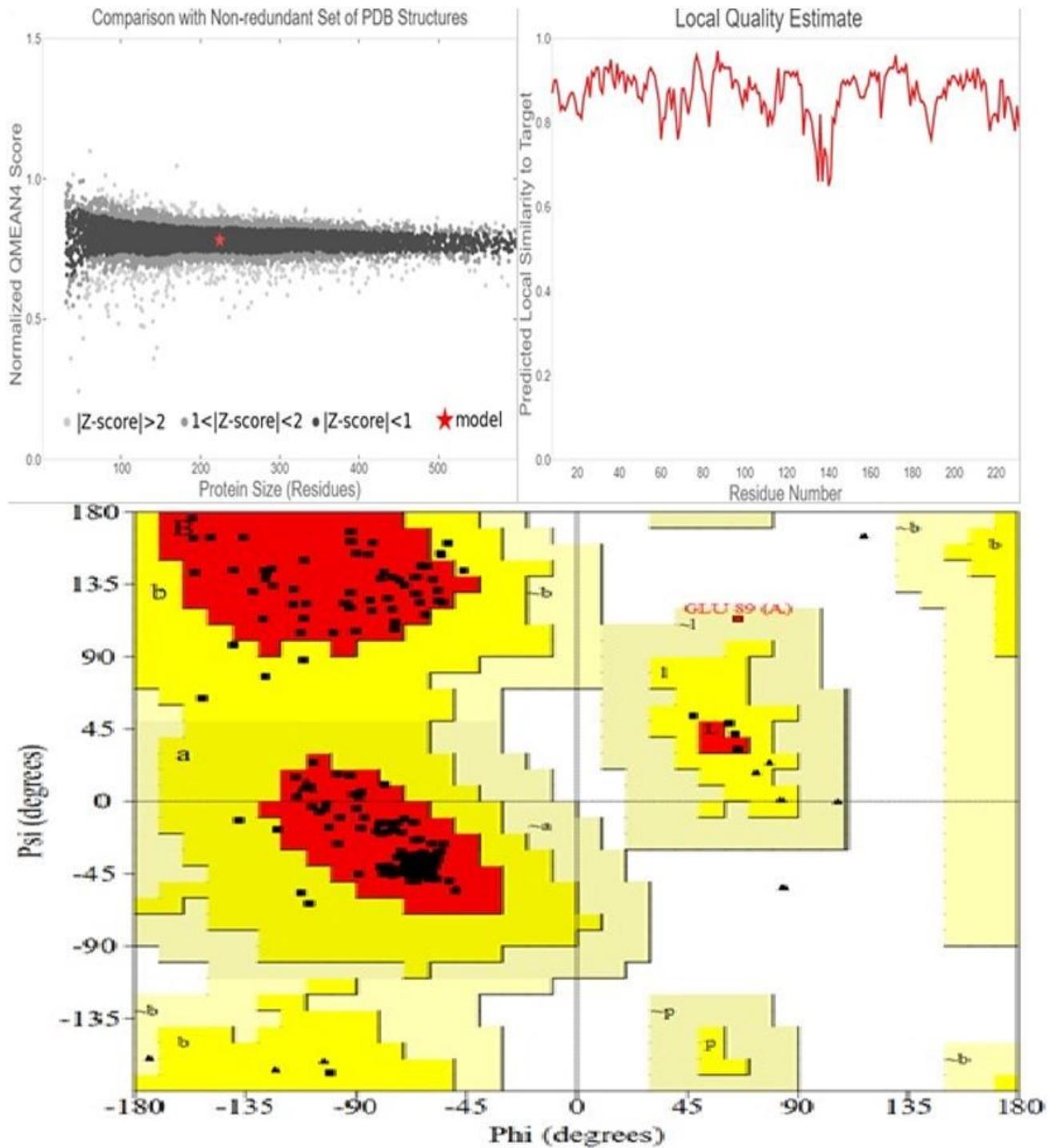


Figure 3

Molecular modelling and validation of selected allergen. A. Quality check of the template used. B. Ramachandra plot generated for the predicted model.

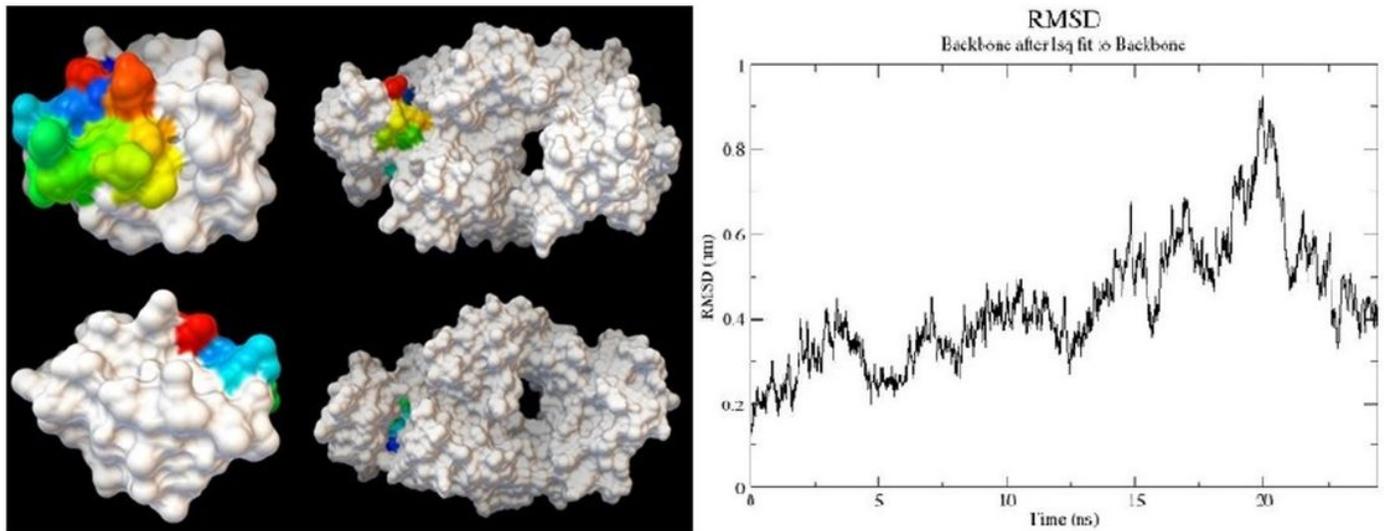


Figure 4

A). Docked structure with highlighted B-cell epitope of A0A5E4EF81. Rainbow color sites depicts the predicted linear conserved B-cell epitopes. B). RMSD of antigen-antibody complex during the molecular simulation analysis for 30 ns.