CONFORMATIONAL EPITOPE PREDICTION OF BIRCH BETV 1 AND HAZEL COR A1 TOWARDS B-CELLS

ABSTRACT

Aim: Background: By attacking with intense specificity employing receptors synthesized or expressed on the surface of antibodies, B-cells will annihilate pathogenic molecules that enter an organism. The molecular interaction observed between both the antibody residue (known as paratope) engaged in binding and the interacting region of the target antigen (known as epitope) is the means through which the destruction of pathogen molecules is achieved. **Objective:** Our aim was to map the antigenic determinants responsible for IgE binding utilizing in silico modelling, simulation of allergenic proteins, and docking of the peptides against IgE. **Methods:** The antigen sequences cut into peptides are docked against the IgE antibody and those with the highest docking scores are further studied for the bond interactions. The overlapping sequences of the high score peptides are observed in the whole antigen model to predict their position. The residues where the bond interactions were found have also been reported for these overlapping peptide sequences. **Results:** The validation is done by antigen-antibody docking studies to confirm the predicted epitope. 25% of the world population suffers from allergic rhinitis and 15% of them develop asthma. White birch and common hazel allergens, namely Bet v 1 and Cor a 1 are known allergens, but their allergen specificity is not yet characterized. **Conclusion:** Negative binding energies of the studied pollen allergens with IgE confirm their allergenicity. Based on the results of overlapping peptides PF 3,4 and PF 16,17 to play a key role in the allergenic response of White birch and Common hazel.

Keywords: BETV 1, COR A1, Birch, Hazel, IgE, B-cell

INTRODUCTION

In the adaptive immune system, B cells play an essential role in protecting the human body against various pathogenic molecules. Specifically, B cells belong to humoral immunity that is mediated through antibodies. In response to exposure to pathogens, B cells develop antibodies that bind to and neutralize the target. However, pathogens are not identified by B-cells as a whole, but through molecular components known as antigens. The component of an antigen that is detected by the immune system, primarily by B cells or T cells, is the epitope,

also known as the antigenic determinant. The majority of the B cell epitopes are conformational (discontinuous), while the remaining (merely 10 percent) are linear B-cell epitopes (continuous). For immunologists, accurate detection of conformational B-cell epitopes is still a major hurdle **[1]**. The major plant allergens of white birch (*Betulaverrucosa*) and common hazel (Corylus avellana) are the antigens chosen for this analysis. Allergic hypersensitivity (aka allergy) is a reaction that happens in an individual when the same allergen is introduced into that person who has developed IgE antibodies in response to that antigen or allergen priorly **[2]**. The most common causative agents of allergic reactions are pollen grains from plants, foods, bee stings, dust mites, molds, fungal spores, animal epithelia, fur and feathers, animal dander, latex. Studies have indicated that allergic disorders such as anaphylaxis, hay fever, atopic dermatitis, eczema, asthma, and many other respiratory and pulmonary diseases affected about 25% of the world population which are primarily caused by aero-allergens. The major hazel allergen Cor a 1 and major birch pollen allergen Bet v 1 are homologous to each other **[3]**. According to researchers, about 53% of people who are allergic to birch pollen also have cross-reactivity to Cor a 1 as well. People who are allergic to birch pollen (70%) also exhibit hypersensitive responses towards different seeds, fruits, nuts, and roots than those without allergy to birch pollen (19%). Hazel allergy is very widespread across European countries and has also been shown to be the most prevalent source of food allergy mediated by IgE **[4].** The present work focused on the in silico molecular characterization of White birch and common hazel Bet v 1 and Cor a 1 allergen-derived peptides, respectively.

METHODS AND MATERIALS

Retrieval of allergen sequences and antibody structure:

The major pollens from the selected plants with high allergenicity were chosen for the study. *Bet v 1* is the major allergen found in the plant White birch (*Betulaverrucosa*). *Bet v 1* is the primary cause of Type I allergies observed during the early springs and winters, especially in the Northern Hemisphere climate zone. *Cor a 1*, is the main pollen of common hazel (*Corylus avellana*). The root of one of the most common food allergies is present in Hazelnut. UniProt is the source used for the retrieval of the FASTA sequences of the major allergens belonging to a selected plant. In general, immunoglobulin E (IgE) is an instinctive response shown by the human immune system to any foreign material (antigen) that has breached into the body. The IgE antibody structure with zero mutations and having a Fab macromolecule also was retrieved from the PDB database **[5-6]**.

UNDER PEER REVIEW

Overlapping and cutting of peptides:

To determine the epitope present in the allergen, the individual peptides are docked against the whole antibody. For this, the entire allergen sequences that were retrieved from UniProt are cut into short overlapping peptide fragments using Sigma-Aldrich tools [Peptide Library design and calculator tool]. Sigma Aldrich's "Overlapping Peptide Fragment Library" tool is used to chop the peptides with a convenient amino acid gap into the appropriate lengths and can also help calculate hydropathy. This tool of Sigma Aldrich is used to cut the amino acid sequence into short peptides of length 10 and with five overlapping amino acids **[7-8]**.

Protein-Peptide Docking:

Docking is performed to predict how short peptide (or allergen peptide) sequences interact with a protein (or antibody). The tool used for docking is HPEPDOCK (A server for protein-peptide docking). HPEPDOCK conducts a docking check for the binding site attributable to a protein receptor structure and a peptide sequence, allowing the peptide to be completely flexible and predicting the protein-peptide complex structure, beginning from random peptide conformations and locations. Computational docking techniques are used to scan for rotational space between a protein receptor and its peptide-binding partner in all possible binding modes. The entire antibody is docked against each overlapping peptide fragment of the pollen sequence. All potential models for each peptide fragment along with their docking scores are estimated by the HPEPDOCK docking. For further analysis, the peptide model with the highest docking score will be identified and picked for every overlapping peptide fragment that has been docked against the antibody, among all potential models that have been predicted by the server **[9].**

Docking analysis:

The antibody and peptide interactions are studied through Schrodinger's Maestro Software, a dynamic molecular modelling analysis tool that offers a flexible environment with multiple display options including full-featured 3D visualization of simple molecules to complex biomolecules. To be precise, the presence of non-covalent bonds and pi-pi interactions between antibody and antigen peptides are analyzed through the Maestro Viewer. The 3D structures of both the receptor (antibody) from the Protein data bank and the antigen peptide model with the highest docking score according to HPEPDOCK are viewed in the Maestro workspace to identify the bond interactions between them. The type of bond that is formed

UNDER PEER REVIEW

between the atoms, the name & number of the atom, residues, and chains at where the bonds are formed for both Antibody and peptide are noted.**[10]**

Homology Modelling:

Homology modelling works on the concept that evolutionarily related proteins bear a roughly equivalent structure. For the prediction of the secondary structure of the antigen sequence that is retrieved from UniProt, a template is essential. The template selection is done through the BLAST tool. Areas of similarities amongst various biological sequences are detected through BLAST. This software helps in the comparison of protein or nucleotide sequences against database sequences and measures their statistical significance. According to the Blastp results, the template is selected. Due to the very high similarity between both the antigen sequences, the Protein Data Bank(PDB) structure (ID:4a86, Birch major allergen) with 72.33% of similarity was selected as a template for building secondary structure for Hazel major allergen (Cor a 1). The model is built or generated using SWISS-MODEL **[11]**.

Antigen-Antibody Docking Studies:

The complete antigen 3D models of both Birch (Bet v1) and Hazel (Cor a1) are docked against the IgE antibody for validation. The structure of the Birch allergen is retrieved from Protein Data Bank (PDB) (ID: 4A86). Whereas the structure of Hazel allergen is modelled through SWISS-MODEL. Both these structures are docked against IgE antibody whose structure is retrieved from PDB (ID: 2vxq). Docking is carried out through the software ClusPro which is a web-based server that is useful for direct protein-protein docking. From all the models that are predicted by the ClusPro server, the models with high scores for both birch and hazel are considered for validation **[12]**.

Results:

Homology modelling:

Template selection: The target sequence was searched with BLAST against the primary amino acid sequence contained in the SMTL. A total of 108 templates were found. Among them, the template with the highest sequence identity, MAJOR POLLEN ALLERGEN BET V 1-A (PDB ID:4a86) is selected as a template with 72.33% similarity.

UNDER PEER REVIEW

Three-dimensional structures of Birch pollen (Bet v 1) were available in the PDB (4A86). Hazel pollen (Cor a 1) structure was determined using the homology modelling application of the SWISS-MODEL.

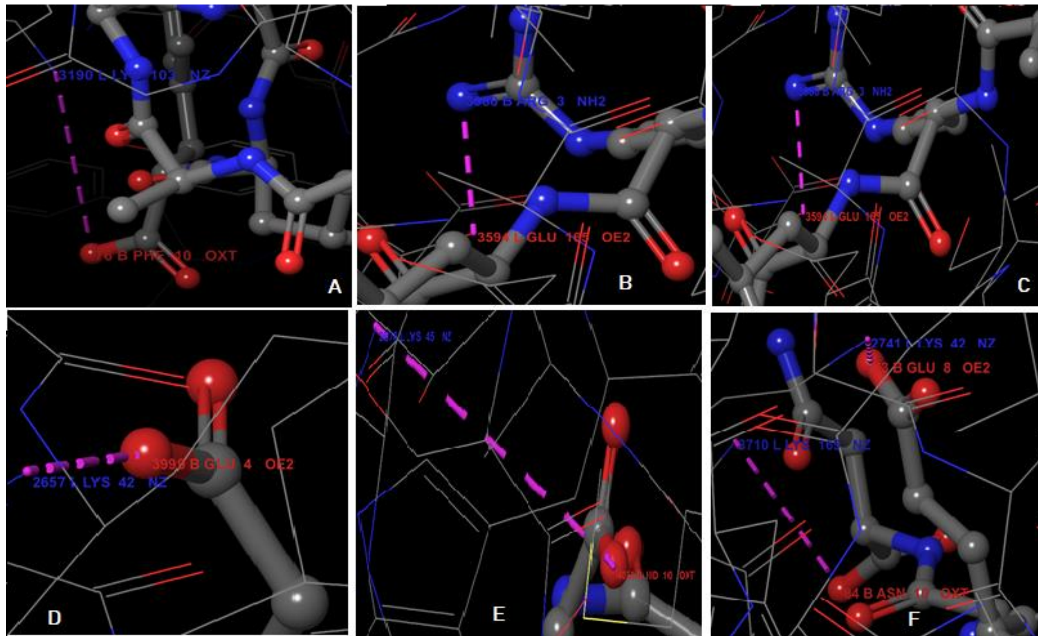

Figure 1: (A-F) Salt Bridge in 3rd peptide of Cor a 1 at residue 'P'. Salt Bridge in 4th peptide of Cor a 1 at residue 'R'. Salt Bridge in 17th peptide of Cor a 1 at residue 'T'. Salt Bridge in 30th peptide of Cor a 1 at residue 'E'. Salt Bridge in 30th peptide of Cor a 1 at residue 'H'. Salt Bridges in 31st peptide of Cor a 1 at residue 'E' and 'N'

UNDER PEER

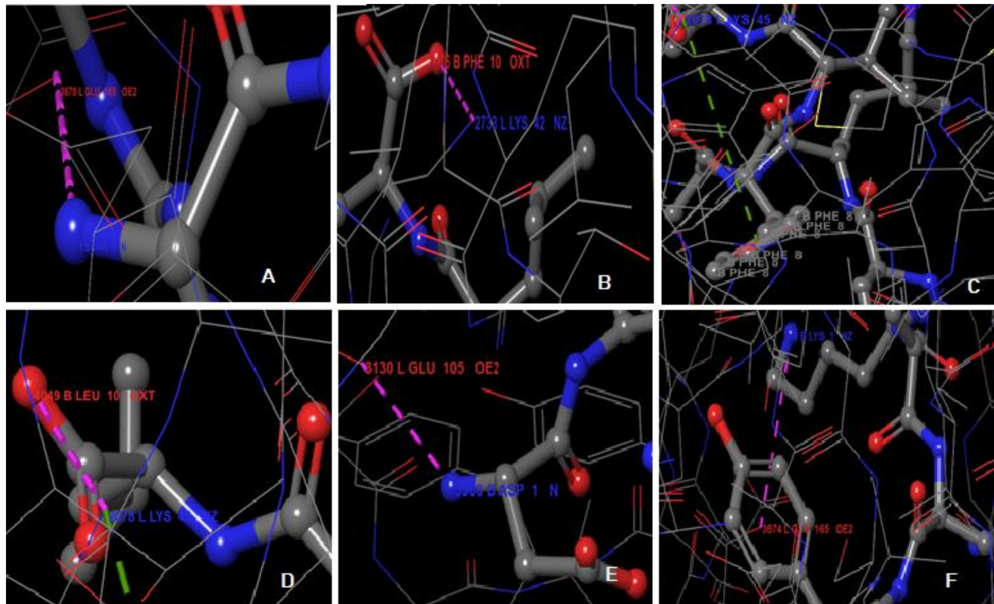

Figure 2: (A-F) Salt Bridge in 3rd peptide of Bet v 1 at residue 'T'. Salt Bridge in 3rd peptide of Bet v 1 at residue 'F'. Pi-cation bond in 4th peptide of Bet v 1 at residue 'F'. Salt Bridge in 4th peptide of Bet v 1 at residue 'L'. Salt Bridge in 16th peptide of Bet v 1 at residue 'D'. Salt Bridge in 17th peptide of Bet v 1 at residue 'K'

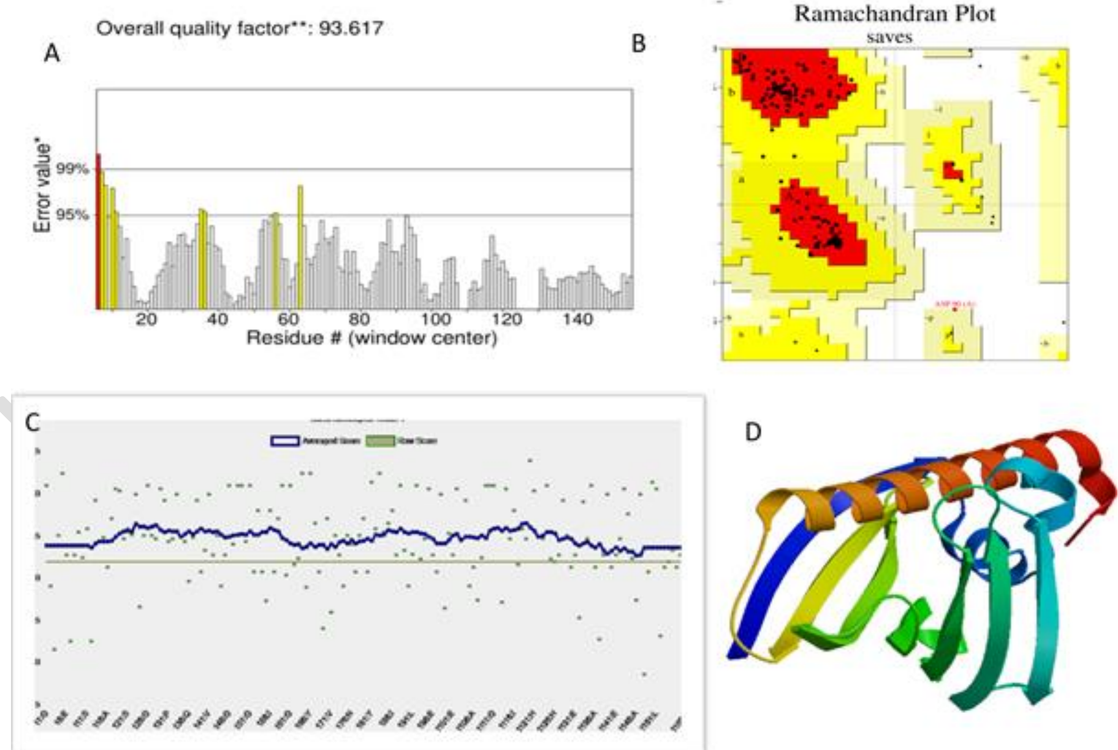

Figure 3: VERIFY 3D analysis indicated that 100% of the residues have averaged 3D-ID score >=0.2. D) 3-D structure of Hazel allergen (Cor a1) modelled through SWISS-MODEL.

Top hit from the Blastp analysis of Hazel pollen (Cor a 1) with PDB ID 4A86 as a template and energy-based model was developed (Figure 3). The structural alignment of the model as evaluated by Ramachandran plot indicated that most of the (93.4%) amino acids fit into the most favored regions, 5.8% of the modelled Cor a 1 residues fall into the additional allowed regions and the remaining were found in generously allowed regions. The ERRAT overall quality factor was 3.617, specifying that the model predicted was good. To identify allergen-IgE interacting sites, an IgE-allergen (protein-protein) docking study was undertaken. IgE antibody (PDB ID: 2VXQ) was retrieved and prepared by using the protein preparation wizard. Concurrently, all the simulated trajectory frames of the modelled allergen of Cor a 1 were clustered based on the energy and deviations. The cluster center frame showing minimal energy, deviations, and fluctuations was chosen for docking studies. Tail-end sequences of the allergen were found intact with the paratope region of the antibody by the end of docking studies. To validate the importance of other amino acids in the allergen, the sequence was divided into overlapping peptides. The allergen sequences of Bet v 1, as well as Cor a 1, were processed using overlapping peptide fragment library software, and 31 different 10-mer peptides, were designed (Bet v 1-31 and Cor a 1-31) with an overlap of five amino acids.

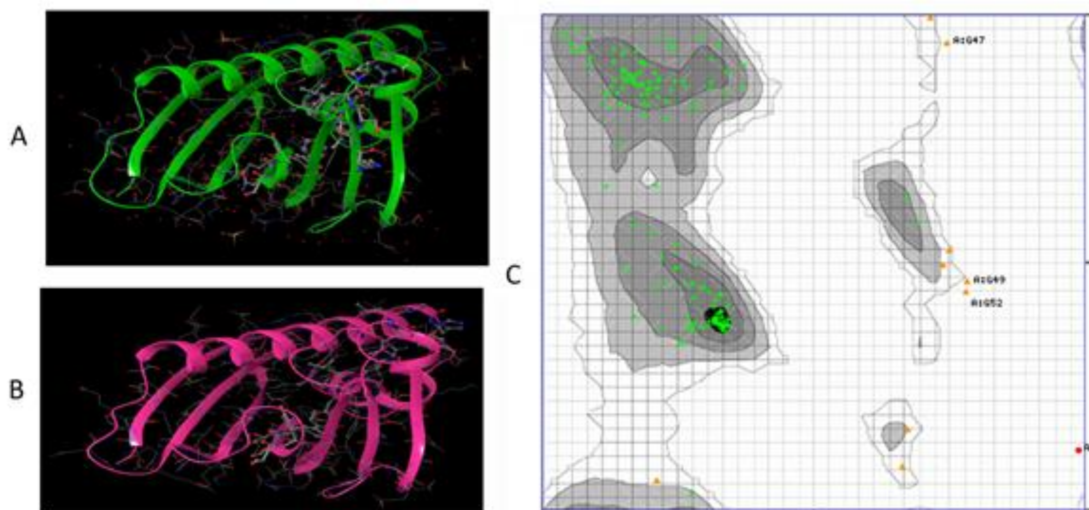

Figure 4: Positions of overlapping peptide fragments in Birch and Hazel allergen.

Initially, the amino acids covering the paratope region of the IgE antibody with respect to bonding interactions with Bet v 1 were identified using the Sitemap module: 5 residues from the heavy chain and 28 residues from the light chain were reported. Similarly, 14 residues from the heavy chain and 31 residues from the light chain of IgE were found to interact with overlapping peptide fragments of Cor a 1. Among the various overlapping peptide fragments

studied for their interactions with IgE, the highest G-scores were revealed for overlapping peptide fragments 3,4; 11,12; and 16,17 of Bet v 1.

On the other hand, overlapping peptide fragments 3, 4; 16, 17; and 30, 31 of Cor a 1 exhibited the highest G-scores. Since both Bet v 1 and Cor a 1 sequences exhibited 74% similarity, it was logical that overlapping peptide fragments 3, 4, and 16, 17 were commonly found to exhibit the highest G-scores for both sequences. Based on the in silico analysis, in Bet v 1 and Cor a 1 overlapping PF 3,4 and 16,17 were identified to have specific IgE paratope interactions and their binding poses are represented in Figure 4.

Discussion:

Bet v 1 is responsible for 60% of allergies with birch (*Betula verrucosa*) pollen released into the air affecting millions of people in spring **[13]**. The birch pollen allergen has different isoforms, all of which exhibit identical conformations, but different allergenic potentials **[14].** IgE and IgG antibodies of patients with allergy to birch pollen serve as tools to define the allergen **[15]**. Up to 90% of the Bet v1-exposed individuals do exhibit IgE-mediated allergic cross-reactions (oral allergy syndrome) to Bet v 1-homologous food allergens, such as hazel nut **[16]**. The three-dimensional structure of Bet v 1 and related pollen and food allergens including Cor a 1 from hazelnut belong to the family of class 10 pathogenesis-related proteins (PR-10) within the Bet v 1 superfamily. PR-10 proteins comprise about 160 amino acid residues with a molecular weight of 17.5 kDa. These proteins exhibit a canonical fold consisting of a seven-stranded antiparallel β-sheet (β1-β7) and three α helices (α1-α3). The two short, consecutive helices α1 and α2 interrupt the β-sheet between strands β1 and β2 while the long C-terminal helix α3 is located above the β-sheet, creating a large and fairly hydrophobic cavity in the protein interior **[17].** Cor a 1 shares 67% sequence identity with Bet v 1 and shared similar tertiary structures based on the homology modelling. As with Cor a 11, structural flexibility in Bet v 1 is distributed across the entire PR-10 scaffold, including secondary structure elements and loops **[18]**. Whether an allergen induces strong immediate-type hypersensitivity reactions in sensitized allergic patients is largely determined by its ability to induce IgE-mediated degranulation of mast cells and basophils **[19]**. The process of degranulation is dependent on cross-linking of cell-bound IgE antibodies and hence requires the presence of at least two IgE epitopes on the allergen **[20].** The IgE antibodies appear to recognize primary conformational epitopes on allergens **[21].** Conformational epitope

mapping using conventional strategies such as testing for IgE reactivity to recombinant or synthetic allergen fragments is not easy because fragmentation of proteins often leads to the loss of the three-dimensional structure of the protein and hence to loss of IgE reactivity **[22].** The onset of birch pollen-related food allergy is believed to be induced by primary sensitization to pollen allergens and subsequent development of secondary food allergy caused by IgE cross-reactivity between homologous pollen and food allergens **[23]**. Bet v 1-specific IgE antibodies were shown to cross-react at the T-cell level with Cor a 1. **[24].** In order to characterize IgE binding as a measure for allergenicity, we characterized the antibody-binding behavior of the Bet v 1 **[25].** IgE recognition of Bet v 1 is not influenced by the bound ligands such as flavonoids **[26]**. We also sought to map the IgE epitopes on the three-dimensional structure of Cor a 1 **[27]**. Due to the lack of crystal structure of Cor a 1, a homology 3D-model was employed for characterizing the epitopes on the surface of Cor a 1 **[28]**. For each of the allergens, namely, Bet v 1 and Cor a 1, and their interaction profiles with Ig E antibodies, the antigen sequence was fragmented into a series of overlapping peptides and their binding modes against IgE were studied. RMSD and RMSF from the simulation results were found to be in the acceptable range of 1-3 A°. The ERRAT score indicates the overall stability of the modelled Cor a 1 protein. Sequential IgE epitope analysis was performed to study IgE epitopes that recognize birch pollen and hazelnut allergens at the level of peptides **[29]**. Our results confirmed a few sequential IgE epitopes, which were found in similar locations and the homology of the amino acid composition of the epitopes of the two allergens was relatively high **[30].** The identified sequential epitopes mapped to the Bet v 1 three-dimensional structures indicate that these residues are exposed on the protein surface and are spread over the β1-α1 regions, β6, β5 and β4 regions in case of Bet v 1 **[31]**. On other hand, in the case of Cor a 1, it involved the β1-α1 regions, β5, β4, and α2 regions. Amino acids 2-11 constitute β1, 113-123 constitute β2, 96-106 constitute β3, 79-87 constitute β4, 68-75 constitute β5, 51-57 constitute β6, 40-45 constitute β7 **[32]**.Similarly, amino acids 15-33 constitute α1, 131-154 constitute α2. The knowledge of the IgE epitopes on the Bet v 1 and Cor a 1 allergens should contribute to the design of effective active and passive immunotherapy strategies for birch pollen and related allergies.

REFERENCES

1. Potocnakova L, Bhide M, Pulzova LB. An introduction to B-cell epitope mapping and in silico epitope prediction. Journal of immunology research. 2016 Oct;2016.

2. Scadding GK, Scadding GW. Innate and adaptive immunity: ILC2 and Th2 cells in upper and lower airway allergic diseases. The Journal of Allergy and Clinical Immunology: In Practice. 2021 Feb 20.
3. Valenta R, Breiteneder H, Pettenburger K, Breitenbach M, Rumpold H, Kraft D, Scheiner O. Homology of the major birch-pollen allergen, Bet v I, with the major pollen allergens of alder, hazel, and hornbeam at the nucleic acid level as determined by cross-hybridization. Journal of allergy and clinical immunology. 1991 Mar 1;87(3):677-82.
4. Schocker F, Lüttkopf D, Müller U, Thomas P, Viehs S, Becker WM. IgE binding to unique hazelnut allergens: identification of non pollen-related and heat-stable hazelnut allergens eliciting severe allergic reactions. European journal of nutrition. 2000 Sep 1;39(4):172-80.
5. Valcour A, Lidholm J, Borres MP, Hamilton RG. Sensitization profiles to hazelnut allergens across the United States. Annals of Allergy, Asthma & Immunology. 2019 Jan 1;122(1):111-6.
6. Dimitrov I, Flower DR, Doytchinova I. AllerTOP-a server for in silico prediction of allergens. InBMC bioinformatics 2013 Apr (Vol. 14, No. 6, pp. 1-9). BioMed Central.
7. Dang HX, Lawrence CB. Allerdictor: fast allergen prediction using text classification techniques. Bioinformatics. 2014 Apr 15;30(8):1120-8.
8. Bateman A, Martin MJ, Orchard S, Magrane M, Agivetova R, Ahmad S, Alpi E, Bowler-Barnett EH, Britto R, Bursteinas B, Bye-A-Jee H. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Research. 2020 Nov 25.
9. UniProt Consortium. Reorganizing the protein space at the Universal Protein Resource (UniProt). Nucleic acids research. 2012 Jan 1;40(D1):D71-5.
10. Wang J, Alekseenko A, Kozakov D, Miao Y. Improved modeling of peptide-protein binding through global docking and accelerated molecular dynamics simulations. Frontiers in molecular biosciences. 2019 Oct 30;6:112.
11. Fiser A. Template-based protein structure modeling. Computational biology. 2010:73-94.
12. Weitzner BD, Jeliazkov JR, Lyskov S, Marze N, Kuroda D, Frick R, Adolf-Bryfogle J, Biswas N, DunbrackJr RL, Gray JJ. Modeling and docking of antibody structures with Rosetta. Nature protocols. 2017 Feb;12(2):401.
13. Asam C, Batista AL, Moraes AH, de Paula VS, Almeida FC, Aglas L, Kitzmüller C, Bohle B, Ebner C, Ferreira F, Wallner M. Bet v 1–a Trojan horse for small ligands boosting allergic sensitization?. Clinical & Experimental Allergy. 2014 Aug;44(8):1083-93.

UNDER PEER REVIEW

14. Vrtala S, Hirtenlehner K, Susani M, Akdis M, Kussebi F, Akdis CA, Blaser K, Hufnagl P, Binder BR, Politou A, Pastore A. Genetic engineering of a hypoallergenic trimer of the major birch pollen allergen, Bet v 1. The FASEB Journal. 2001 Sep;15(11):2045-7.
15. Jarolim E, Rumpold H, Endler AT, Ebner H, Breitenbach M, Scheiner O, Kraft D. IgE and IgG antibodies of patients with allergy to birch pollen as tools to define the allergen profile of Betulaverrucosa. Allergy. 1989 Aug;44(6):385-95.
16. Herkenrath C, Gottmann-Lückerath I, Steigleder GK. Combined allergy against hazel pollen and hazel nuts. Zeitschrift fur Hautkrankheiten. 1982 Oct 1;57(19):1399-405
17. Fernandes H, Michalska K, Sikorski M, Jaskolski M. Structural and functional aspects of PR-10 proteins. The FEBS journal. 2013 Mar 1;280(5):1169-99.
18. Fernández-Quintero ML, Loeffler JR, Waibl F, Kamenik AS, Hofer F, Liedl KR. Conformational selection of allergen-antibody complexes—surface plasticity of paratopes and epitopes. Protein Engineering, Design and Selection. 2019 Nov;32(11):513-23.
19. Valenta R, Kraft D. Recombinant allergen molecules: tools to study effector cell activation. Immunological reviews. 2001 Feb;179(1):119-27.
20. Segal DM, Taurog JD, Metzger H. Dimeric immunoglobulin E serves as a unit signal for mast cell degranulation. Proceedings of the National Academy of Sciences. 1977 Jul 1;74(7):2993-7.
21. Gadermaier E, Flicker S, Aberer W, Egger C, Reider N, Focke M, Vrtala S, Kundi M, Valenta R. Analysis of the antibody responses induced by subcutaneous injection immunotherapy with birch and Fagales pollen extracts adsorbed onto aluminum hydroxide. International archives of allergy and immunology. 2010;151(1):17-27.
22. Vrtala S, Hirtenlehner K, Vangelista L, Pastore A, Eichler HG, Sperr WR, Valent P, Ebner C, Kraft D, Valenta R. Conversion of the major birch pollen allergen, Bet v 1, into two nonanaphylactic T cell epitope-containing fragments: candidates for a novel form of specific immunotherapy. The Journal of clinical investigation. 1997 Apr 1;99(7):1673-81.
23. Irañeta SG, Seoane MA, Laucella SA, Apicella C, Alonso A, Duschak VG. Antigenicity and immunocrossreactivity of orange tree pollen and orange fruit allergenic extracts. International archives of allergy and immunology. 2005;137(4):265-72.
24. Jahn-Schmid B, Radakovics A, Lüttkopf D, Scheurer S, Vieths S, Ebner C, Bohle B. Bet v 1142-156 is the dominant T-cell epitope of the major birch pollen allergen and important for cross-reactivity with Bet v 1–related food allergens. Journal of allergy and

UNDER PEER REVIEW

clinical immunology. 2005 Jul 1;116(1):213-9.

25. Spangfort MD, Mirza O, Ipsen H, Van Neerven RJ, Gajhede M, Larsen JN. Dominating IgE-binding epitope of Bet v 1, the major allergen of birch pollen, characterized by X-ray crystallography and site-directed mutagenesis. The Journal of Immunology. 2003 Sep 15;171(6):3084-90.
26. Kofler S, Asam C, Eckhard U, Wallner M, Ferreira F, Brandstetter H. Crystallographically mapped ligand binding differs in high and low IgE binding isoforms of birch pollen allergen bet v 1. Journal of molecular biology. 2012 Sep 7;422(1):109-23.
27. Negi SS, Braun W. Cross-React: a new structural bioinformatics method for predicting allergen cross-reactivity. Bioinformatics. 2017 Apr 1;33(7):1014-20.
28. Vyas VK, Ukawala RD, Ghate M, Chintha C. Homology modeling a fast tool for drug discovery: current perspectives. Indian journal of pharmaceutical sciences. 2012 Jan;74(1):1.
29. Dall'Antonia F, Pavkov-Keller T, Zangger K, Keller W. Structure of allergens and structure based epitope predictions. Methods. 2014 Mar 1;66(1):3-21.
30. McClain S. Bioinformatic screening and detection of allergen cross-reactive IgE-binding epitopes. Molecular nutrition & food research. 2017 Aug;61(8):1600676.
31. Ahammer L, Grutsch S, Kamenik AS, Liedl KR, Tollinger M. Structure of the major apple allergen Mal d 1. Journal of agricultural and food chemistry. 2017 Mar 1;65(8):1606-12.
32. Lotta LA, Scott RA, Sharp SJ, Burgess S, Luan JA, Tillin T, Schmidt AF, Imamura F, Stewart ID, Perry JR, Marney L. Genetic predisposition to an impaired metabolism of the branched-chain amino acids and risk of type 2 diabetes: a Mendelianrandomisation analysis. PLoS medicine. 2016 Nov 29;13(11):e1002179.

UNDER PEER REVIEW