Computational characterization and epitope prediction for Bet-v1 like protein of *Cannabis sativa*

Zarrin Basharat *

Department of Environmental Sciences
Fatima Jinnah Women University
46000, Pakistan

**Corresponding author*
E-mail: zarrin.iiui@gmail.com

Abstract: *Cannabis sativa encodes a Bet-v1 like protein is an allergen and a causuative agent of pollen allergy. Multiple sequence alignment of this protein revealed conserved residues in Betv1 domain. Identification of linear epitopes of this protein was done after preliminary bioinformatics characterization and structure prediction. Structure prediction was done using Modeller software and minimized using Swiss PDBViewer. Six linear epitopes were then, predicted using EMBOSS antigenic program. Phylogenetic analysis of Bet-v1 with other sequences demonstrated divergence patterns with allergens of other species but revealed conserved residues in allergenic epitopes. This study can serve as an informational aid in the development of hypoallergenic vaccine for Cannabis sativa allergy.*

Keywords: *Cannabis sativa, allergen, epitope prediction, Betv1 like protein*

Introduction

Hemp (*Cannabis sativa)* is a weed specie known to cause allergy for more than six decades [9]. Hemp pollen exposure is injurious to respiratory function [14]. Betv1 like protein in this weed is a biomedically important protein as it is a potent aeroallergen in areas with temperate climates. Analysis of molecular size and allergen content are functional approach for allergoid classification and characterization [3]. It is difficult to infer characteristics distinguishing this allergen from non-allergen protein sequences manually so high throughput omic-based approaches are harnessed for studying regular characteristics of allergens. Currently a great deal of effort is being put in the development of innovative, swift and accurate epitope mapping tactics as these can aid in the identification of epitopes and regulating chemicals of allergen vaccines for IgE binding reduction, resulting in improvement of vaccines by boosting safety and at the same time maintaining clinical efficacy. Protein drug immunogenicity is a noteworthy venture in the therapeutics formulation procedure. For this purpose, bioinformatics based studies were carried out on this protein in order to gain a better understanding of its allergenicity. Immunological action of antibody is dependent upon explicit binding on a distinct sites on the targetted antigen called as epitopes. This tactic can in addition be used for spotting attachment sites and interfaces of diverse protein forms involving interactions even apart from the immunological milieu [10].The present investigative study focuses on characterization of Betv1 and sequence as well as structure based allergenic linear epitope to aid in the identification of functional amino acid sites to assist vaccine development.

Materials and methods

Sequence Retrieval

The primary sequence of Betv1 like allergen protein was acquired from the NCBI database with accession number AFN42528. The data was used for structure prediction and allergenicity assessment.

Physicochemical analysis, structure and phylogenetic analysis

Physiochemical analysis of the sequences were done using PROTPARAM tool [5,13].Subcellular localization was checked using CELLO v.2.5 (http://cello.life.nctu.edu.tw/) and phosphorylation profile study was carried out using NetPhos 2.0 server (http://www.cbs.dtu.dk/services/NetPhos/). Modeller v9.1 [4] was used to predict structure of allergen of *Cannabis sativa* based on the structure of major cherry allergen PRU AV1 with Protein Data Bank ID:1E09 to get the .pdb file for epitope prediction and mapping on protein surface. Qmean server [2] was used for quality assessment and Rampage analysis [8] was done for structure validation. The protein structure was further energy minimized using Swiss PDBViewer (spdbv.vital-it.ch/). Phylogenetic analysis was carried out to study evolutionary pattern using software MEGA v5.2 [12]. Output files in .png and .pdf format were obtained.

Allergenic epitope detection

Allergenic domains of the sequences were extracted by scanning against Pfam database [1]. EMBOSS antigenic program (http://emboss.bioinformatics.nl/cgi-bin/emboss/antigenic) revealed antigenic sites on the query protein.

Results and discussion

Epitope detection utilizing bioinformatics techniques has become prevalent. Standardization and optimization of such *in silico* approaches can be exploited for potent antigenic site discovery. The role of highly conserved residues plays a major role in determining the allergenicity as these can then, be used for targeting allergenic proteins to pave way for vaccine development. Multiple sequences alignment was done for studying conservation among residues using ClustalW and phylogenetic analysis utilizing Neighbour joining method (Figure 1).

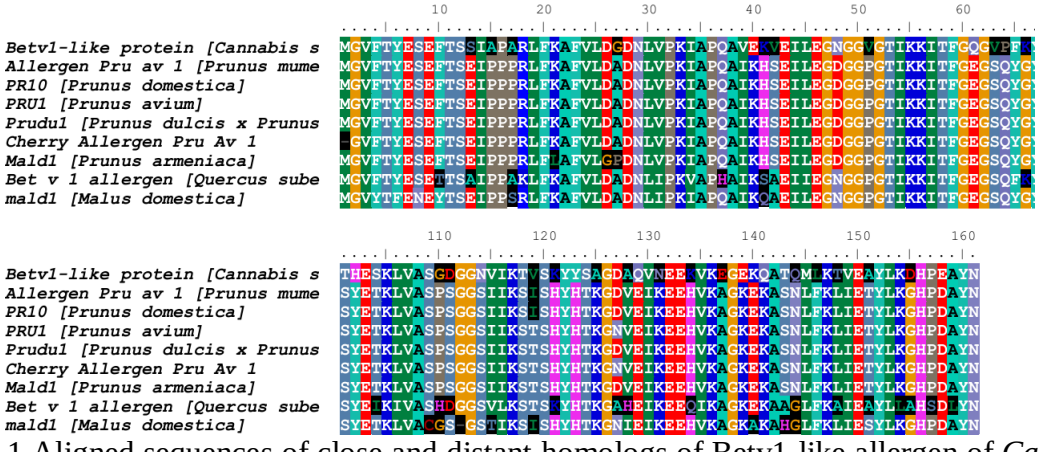

Figure 1 Aligned sequences of close and distant homologs of Betv1-like allergen of *Cannabis sativa*

Alignment revealed that pathogenesis related protein of *Prunus mume*, *Prunus domestica*, *prunus avium* are close homologs while Betv1 allergen of *Quercus suber* and Mal-d1 like allergen proteins of *Malus domestica* were found to be remote homologs.

Primary analysis of this allergen revealed that it is a 17607.9 dalton soluble protein with an average hydrophobicity of -0.364596 and instability index of 26.61 representing stability of the protein. Aliphatic index was 84.16 and theoretical pI was estimated to be 5.29. CELLO results indicated that the protein was cytoplasmic in nature. Neural network based analysis revealed that the allergen was phosphorylated at 53, 67, 85 and 124 amino acid positions (Figure 2).

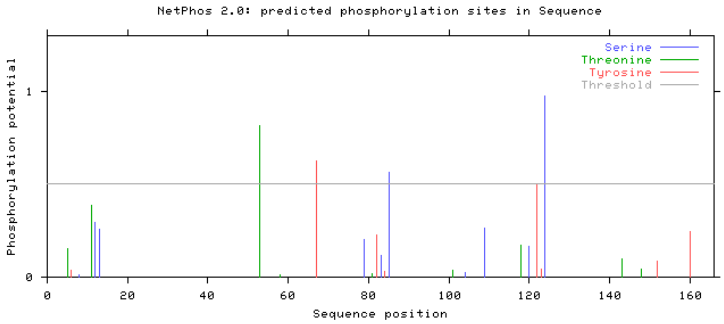

Figure 2 Predicted phosphorylation sites in the Betv1-like protein using NetPhos 2.0 server.

Pfam database analysis revealed presence of a single Betv1 domain from 1 to 154 position of amino acids in the sequence. 3D structure was modelled (Figure 3) using cherry allergen structure template, energy minimized and validated after elementary sequence analysis.

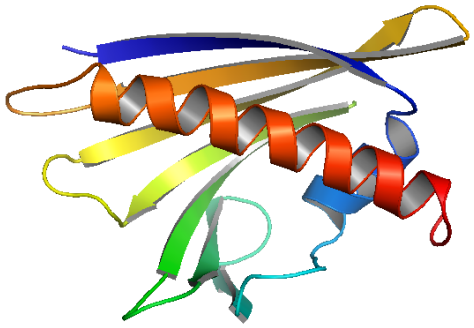

Figure 3. 3D structure of the Betv1-like protein.

Qmean score of 0.6 was obtained and was agreeable as it was between the anticipated value 0-1. Rampage analysis revealed that 91.8% residues lied in the favoured regions, 7.6% in the allowed region and only 0.6% in the outlier region. As more than 90% residues lied in favoured region so the model was considered satisfactory.

The understanding of protein antigenicity and interaction with the immune system is of great interest to improve the control of diseases. Since the linear antigenic map of Betv1 remained almost unexplored, linear epitopes were studied. Six linear epitopes were predicted using antigenic program from EMBOSS (Table 1).

Table 1. Predicted linear Epitopes from Emboss Antigenic program

No.	Predicted by	Start position	End position	peptide	Number of residues	score
1	Antigenic	11	45	TSSIAPARLFKAFVLDGDNLVPKIAPQAVEKVEIL	35	1.151
2	Antigenic	60	73	GQGVPFKYVKHKIE	14	1.135
3	Antigenic	116	128	IKTVSKYYSAGDA	13	1.103
4	Antigenic	80	87	LTYSYSII	8	1.094
5	Antigenic	103	109	ESKLVAS	7	1.085
6	Antigenic	146	157	LKTVEAYLKDHP	12	1.078

Linear epitopes can be more defiant to harm than conformational epitopes and their conservation in different proteins can act as a tailored tool for antigen discovery. Antigenic sites portrayal has been accomplished for some proteins but attention has been almost exclusively focused on epitopes involved in neutralization. The acquired antigenic region in the present study can be exploited for designing efficient vaccine against aeroallergen posed by *Cannabis sativa*. Linear epitope mapping on functional sites can lead to improved antigenic structural perception of Bet-v1 and related allergen proteins. Related previously documented works imply that allergens have a propensity to share specific sequence similarities [7, 11]. Besides clinical and lab testing, present methods of allergenicity inference entail an initial evaluation of the allergen protein sequence with homologues having conserved residues [6]. These can also be used to establish evolutionary relationships. Consequently, sequence characterization can lead to prediction of prospective allergenicity of proteins.

Conclusion

This study is an attempt to characterize the allergen of Cannabis sativa in terms of structure and allergenicity. It is a baseline informational aid to the wealth of existing knowledge pertaining to allergen structures and epitope analysis and can be further utilized for epitope mapping expending reverse vaccinology approach. It can also serve as an informational aid for epitope analysis with monoclonal antibody testing in the lab.

References

1. Bateman A., E. Birney, L. Cerruti, R. Durbin, L. Etwiller, S.R. Eddy, S. Griffiths-Jones, K.L. Howe, M. Marshall, E. L.L. Sonnhammer (2002). The Pfam Protein Families Database. Nucleic Acids Research,30,276-280.
2. Benkert P., M. Kunzli, T. Schwede (2009) . QMEAN server for protein model quality estimation.Nucleic Acids Research, 37, W510-4.
3. Carnés J., M. Himly, M. Gallego, V. Iraola, D.S. Robinson, E. Fernández-Caldas, P. Briza (2009). Detection of allergen composition and in vivo immunogenicity of

depigmented allergoids of *Betula alba*. Clinical Experiental Allergy, 39(3),426-434.

4. Eswar N., M.A. Marti-Renom, B.M.S. Webb, D. Madhusudhan, M. Eramian, U. Shen, U. Pieper, A. Sali (2006). Comparative Protein Structure Modeling With MODELLER. Current Protocols in Bioinformatics, John Wiley and Sons Inc. Supplement 15, 5.6.1-5.6.30.
5. Gasteiger E., C. Hoogland , A. Gattiker, S. Duvaud, M.R. Wilkins, R.D. Appel, A. Bairoch (2005). The Proteomics Protocols Handbook, Humana Press.
6. Kong W., T.S. Tan, L. Tham, K.W. Choo (2006). Improved prediction of allergenicity by combination of multiple sequence motifs.In Silico Biology, 7(1), 77 -86.
7. Ladics G.S., M.K. Selgrade (2009). Identifying food proteins with allergenic potential: evolution of approaches to safety assessment and research to provide additional tools. Regulatory Toxicology and Pharmacology, 54,S2.
8. Lovell S.C., I.W. Davis, W.B. Arendall III, P.I.W. deBakker , J.M. Word, M.G. Prisant, J.S. Richardson, D.C. Richardson (2002). Structure validation by Calpha geometry: phi,psi and Cbeta deviation. Proteins: Structure, Function and Genetics, 50, 437-450.
9. Mayoral M., H. Calderón, R. Cano, M. Lombardero (2008). Allergic rhinoconjunctivitis caused by *Cannabis sativa* pollen. Journal of Investigive Allergology and Clinical Immunology, 18(1),73-4.
10. Mayrose I., T. Shlomi , N.D. Rubinstein, J.M. Gershoni, E. Ruppin, R. Sharan, T. Pupko (2007). A graph-based algorithm for epitope mapping using combinatorial phage-display libraries. Nucleic Acid Research, 35(1),69-78.
11. Sharma R., A.K. Singh, V. Umashankar (2009). Characterization of allergenic epitopes of Ory s1 protein from *Oryza sativa* and its homologs. Bioinformation, 4(1),12-18.
12. Tamura K., D. Peterson, N. Peterson, G. Stecher, M. Nei, S. Kumar (2011**).** MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology and Evolution, 28, 2731-2739.
13. Wilkins M.R., E. Gasteiger, A. Bairoch, J.C. Sanchez, K.L. Williams, R.D. Appel, D.F. Hochstrasser (1999). Protein identification and analysis tools in the ExPASy server. Methods in Molecular Biology, 112,531-52.

14. Zuskin E., L. Mustajbegovic, E.N. Schachter (1994). Follow-up study of respiratory function in hemp workers. American Journal of Industrial Medicine, 26,103-15.