Computational identification of *Brassica napus* pollen specific protein Bnm1 as an allergen

Md. Rezaul Islam[a,*], Ahsan Habib Polash[b], M Sadman Sakib[a], Chinmoy Saha[c],

Atiqur Rahman[a]

[a]Department of Biochemistry and Molecular Biology, University of Dhaka, Dhaka-1000, Bangladesh

[b]Molecular Biology Laboratory, Department of Biochemistry and Molecular Biology, University of Dhaka, Dhaka-1000, Bangladesh

[c]Postgraduate School of Molecular Medicine, Erasmus MC, Rotterdam, The Netherlands.

E-mail address: rezaul.nayeem@gmail.com

Abstract

Bnm1 is a pollen specific protein from Brassica napus (oilseed rape) and it is specifically expressed in the bi-cellular and tri-cellular stages of pollen development. Since the incidence of pollinosis due to oilseed rape (Brassica napus) is increasing day by day, keeping pace with its high cultivation rate, the search for its allergens is a demand of time to develop effective immune therapy. In the present study, different computational tools were adopted to predict the potential of Bnm1 as a candidate allergen. Physico-chemical properties of Bnm1 showed its molecular weight (~20kD) and theoretical pI (5.27) along with other properties to be fallen between the ranges essential for a protein to be an allergen. Keeping in mind the capability of allergen to induce both humoral and cell mediated immune response, we checked both the potential B cell and T cell epitope candidates of Bnm1 using different immune-informatics tools housed at IEDB analysis resource. For B cell epitope prediction, potential antigenic sites on the protein surface were predicted by both propensity scale and machine learning method followed by their mapping on Bnm1 3D structure predicted from homology modeling. In case of T cell epitope prediction, interaction of the core sequence with seven abundant MHC-II alleles (DRB1*0101, DRB1*0301, DRB1*0401, DRB1*0701, DRB1*1101, DRB1*1301 and DRB1*1501) within an IC50 range (IC50<25) was the basis. We observed multiple epitope interactions with DRB1*0101 allele, implying that this interactions might evoke a strong TH2 response ensuring an increased production of IgE response. Based on our prediction we hereby claim that the Bnm1 is a potential allergen having capability to induce both humoral and cell mediated allergic responses. However, in vitro analysis was out of our scope and therefore it should have to be performed to validate the potency of Bnm1 as an allergen.

Key words: *Bnm1, Brassica napus, allergen, pollinosis, in silico, epitope*

1. Introduction

Allergens are mostly small proteins or protein bound small substances having a molecular weight ranging from 15 to 40 kD [1]. These proteins or protein associated substances not only can induce IgE antibody mediated immune response but also can induce other antibody responses like IgA, IgM, and IgG [2]. Moreover, they can elicit T cell response in the human body system which is mostly Th2 cell mediated [3, 4]. The sources of these allergens can be different food items, foreign serums, insect venoms, pollens, insect products etc. [5]. Most of the allergens have biological function but little is known about their functional relevance to allergenicity. The functional property of allergen like enzyme activity may contribute to the induction of allergenicity [6]. Pollen allergy (pollinosis) has brought a large attention to the scientific

community around the world and the rate of pollinosis is increasing day by day [7, 8]. Among the symptoms of immediate response upon exposure to pollen allergens are allergic rhinitis, asthma and hay-fever [9-11] etc. Individuals who are hypersensitive to pollen often show allergic like symptoms.

Brassica napus (oilseed rape), belonging to cruciferae family, is one of the most cultivated crop plants around the world. The importance of this plant as a source of oil and medicinal components is reviewed elsewhere. Cultivation rate of oilseed rape around the globe has been increasing over the last 10 years resulting in about 31.5 Mha of lands used to cultivate in 2010 [12] The recent introduction of genetically engineered oil seed has geared the cultivation rate even more [13, 14].

With the pace of the cultivation rate of this plant an alarming concern for allergenicity is generated, since pollen from oilseed rape have been reported as potential allergen. In the most abundant regions, most patients with hay-fever blame the blooming of *Brassica napus* for their symptoms during its flowering season [15]. During this flowering season the microscopic male cells of the plant make round or oval structural grain which is programmed to transport along the wind to help in the process of pollination. Pollens of *Brassica napus* guided by the wind are spreaded everywhere around the cultivated land and people mostly farmers having exposed to it suffer from hay-fever, allergic rhinitis like seasonal diseases [15]

Although some of the high molecular mass proteins in oilseed rape pollen have been identified as allergens [16-20], a quest for other allergens is going on in different laboratories. In modern biology, *in silico* approaches have made it possible to give a straightforward time and money saving way to find out the solution of many biological problems like vaccine design [21], prediction of deleterious effects of mutations [22] and many more. It can also predict potential allergens from a given whole proteome.

Bnm1 is a pollen specific protein in *Brassica napus* and it is specifically expressed in the bicellular and tricellular stages of pollen development [23]. In the present study we have attempted to analyze the potentials of Bnm1 to be an allergen using different computational tools and immune-informatics databases. To our knowledge, this is the first immune-informatics approach to examine the potency of Bnm1 from oilseed rape to be an allergen.

2. Materials and Methods:

2.1 Protein Sequence retrieval

Bnm1 (pollen specific protein) protein sequence (P93760) for *Brassica napus* was retrieved from the UniProt database (http://www.uniprot.org/). This protein sequence was the basis to perform different computational predictions from linear amino acid residues.

2.2 Prediction of physico-chemical properties

Different physico-chemical properties for Bnm1 protein was predicted from its linear amino acid sequence. ProtParam tool was employed to predict various physical and chemical properties for Bnm1 protein including the molecular weight, theoretical pI, atomic composition, amino acid composition, instability index, extinction coefficient, grand average of hydropathicity (GRAVY), estimated half-life, and aliphatic index [24].

2.3 Potential antigenic sites prediction

Antigenicity of the protein was predicted by determination of hydrophobic-hydrophilic region and thereby determining the regions of protein, exposed to outer surface and can react to B cell. To predict the potential antigenic sites two prediction methods called Kolaskar-Tongaonkar antigenicity and Parker's hydrophilicity were assigned followed by determination of antigenic propensity and hydrophilicity respectively from the plots generated [25, 26].

2.4 Potential B cell epitope prediction:

All the regions exposed to outer surface are not potent site to react with B cell, that's why B cell epitopes are needed to be predicted. A machine learning tool BepiPred housed at IEDB analysis tools was assigned to predict B cell epitope [27]. This tool uses a combinatorial algorithm comprising both hidden markov model and antigenic propensity and thus was used to cross check the predicted result from Parker's hydrophilicity and Kolaskar-Tongaonkar antigenicity prediction method [25, 26].

2.5 Prediction of 3D structure of Bnm1 and mapping of B cell epitopes on the structure

For 3-D structure prediction we followed a template based Bnm1 protein structure modeling using RaptorX (http://raptorx.uchicago.edu/) server which makes several alignments of the target protein sequence with different protein templates having sparse protein sequence to predict the *ab initio* model of protein, which is essential to have a better structure prediction [28]. Based on the alignment score top ten ranked alignments were predicted by probabilistic-consistency algorithm and a novel nonlinear scoring function. We chose the structure of Bnm1 which showed maximum alignment score between the Bnm1 protein and its target template (RCSB PDB ID 1x8z, chain A). Energy minimization of the structure was performed by Swiss PDB viewer tool [29]. To validate the structure a Ramachandran plot was generated using ProCheck program [30] which measures the stereo-chemical properties of the protein structure.

2.6 Potential T cell epitope prediction

To check whether the Bnm1 protein can elicit T cell response, we predicted potential T cell epitopes using "Peptide binding to MHC class II molecules" program under "MHCII binding prediction" tool in IEDB analysis resource. For our prediction we followed NetMHCIIpan prediction method choosing the seven abundant HLA class II alleles DRB1*0101, DRB1*0301, DRB1*0401, DRB1*0701, DRB1*1101, DRB1*1301 and DRB1*1501 from the selection panel [31, 32]. Later, from the predicted T cell epitopes only those epitopes having IC50 score less than 25 were selected as potential T cell epitope candidates since epitopes having IC50 value < 25 shows higher binding affinity with MHC alleles [33].

3. Results and Discussion

3.1 Physico-chemical properties predicts similar characteristics of Bnm1 like conventional allergen

Physico-chemical properties of an intact protein sometimes can predict the allergenic potential of a protein [34]. Bnm1 consists of 182 amino acids and the molecular weight was predicted to be approximately 20 kD (Table 1). The total amino acid distribution along the Bnm1 protein (Figure 1) shows that there is no tryptophan residue present and among the 20 amino acids six amino acid residues e.g. Ala, Asp, Leu, Ser, Thr and Val comprise almost 60% of total composition of Bnm1 protein (Figure 1). Since all of these highly abundant residues have acidic pI (minimum 2.77 for Asp to maximum 6.0 for Ala) range, this suggests the protein's theoretical pI to be acidic which is a combinatorial score of all the pI scores for all the amino acid residues present in Bnm1 protein.

Table 1 Detail of physico-chemical properties for Bnm1 protein from oilseed rape

| Properties	score
Number of amino acids	182
Molecular weight	19651.2
Theoretical pI	5.27
Total number of negatively charged residues (Asp + Glu)	24
Total number of positively charged residues (Arg + Lys)	17
Total number of atoms	2759
Ext. coefficient($M^{-1}cm^{-1}$)	7575
The estimated half-life	30 hours
The instability index (II)	40.27
Aliphatic index	88.63
Grand average of hydropathicity (GRAVY)	-0.128

ProtParam prediction shows that the protein is acidic (pI 5.27) and net negatively charged as total number of negatively charged residues outnumber the total number of positively charged residues (Table 1), hence it may preferentially be processed by dendritic cells [34]. Based on the half lives of the individual N-terminal amino acid residues, the predicted overall half life of the Bnm1 [35, 36] was quite a high as 30 hours. The predicted instability index of the Bnm1 indicates that this protein is slightly unstable. The grand average of hydropathicity predicted from the Bnm1 linear protein sequence was predicted as negative and hence most of the amino acid residues in Bnm1 protein are likely to be present on the surface of the folded Bnm1 protein.
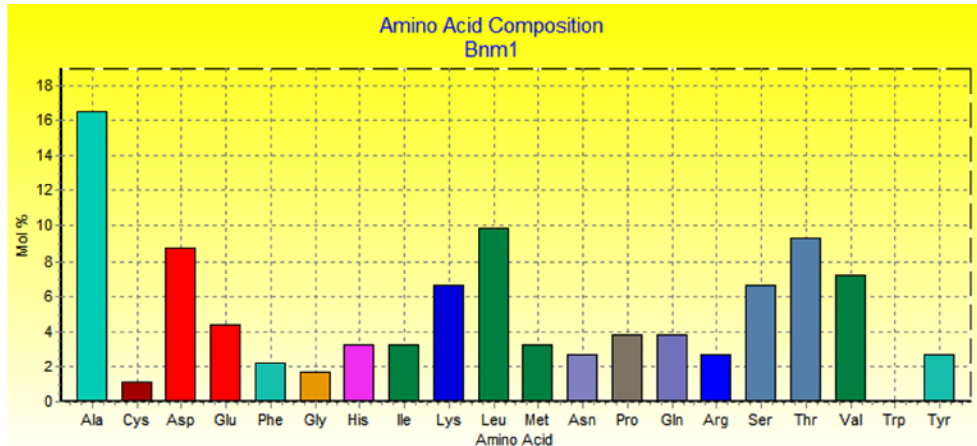
Figure 1 Amino acid composition of Bnm1 Among the all amino acids present in Bnm1, Alanine (Ala) is ~16.5% of total composition and thus outnumbers the other amino acid. Cysteine (Cys) is of least percentage present in Bnm1 protien. Note there is no tryptophan residue in its total composition.

3.2 Potential antigenic sites are on the surface of the Bnm1 protein

To show allergenic response the whole protein does not need to be antigenic; rather there are some antigenic determinants which are called epitope accounts for the immunogenic reactions. It is hypothesized that the antigenic epitope must reside in the outer layer or the hydrophilic region of the protein so that it can interact with others. So simply hydrophobic or hydrophilic profile of a protein sometimes can predict a protein's potential as an allergen from the linear amino acid sequence [37]. For prediction of Bnm1 allergenicity, Kolaskar and Tongaonkar prediction method was employed which functions on the basis of physico-chemical properties of amino acids in proteins and abundances in experimentally known epitopes [25]. In Kolaskar scale X axis represents amino acid residues whereas Y axis represents anigenic propensity of the protein.
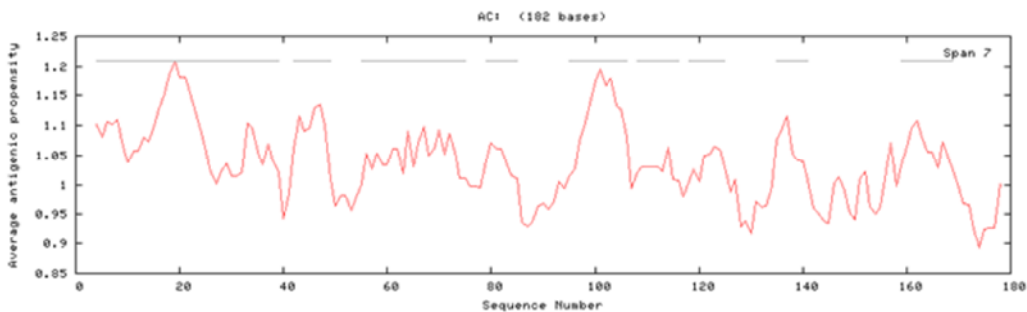

Figure 2 Kolaskar and Tongaonkar antigneicity graphical plot. The protein sequences those satisfied the set threshold value (antigenic propensity threshold 1.00) are predicted to be potential antigenic site against which antibodies can elicit the response.

Table 2 Potential antigenic peptide sequences predicted from kolaskar and Togaonkar antigenicity plot

number	Start position	End position	Peptide sequence	Peptide length
1	4	40	FSVLSTFAAAAITLQLLLVPASASPHM KYIDAICDRS	37
2	43	50	QDYCVKTL	8
3	56	76	TAAPIGLNPLAEVMALTIAHA	21
4	80	86	AAFAET	7
5	96	107	YHKAYLAWADL	12
6	109	117	SANLKLKQS	9
7	119	126	DTAHYDVR	8
8	136	142	EGLVASK	7
9	160	170	LLDLAASAADA	11

For Bnm1 the average antigenic propensity of the bnm1 protein is 1.037 so all residues having a value greater that 1.037 are potential antigenic determinant (Figure 2). Nine peptides are found to satisfy the threshold (1.00) value set prior to analysis and thus are potential antigenic sites of the protein and they have the potential to evoke B cell response. The detail of the individual peptide is summarized in Table 2 .The peptide regions ranging from 4 to 39 amino acid residue having "FSVLSTFAAAAITLQLLLVPASASPHMKYIDAICDR" sequence and 55 to 75 amino acid residue having "PTAAPIGLNPLAEVMALTIAH" sequence are predicted to have the highest antigenic propensity score and both in total comprise about 38% portion as allergenic sites of all the sites in Bnm1 protein.
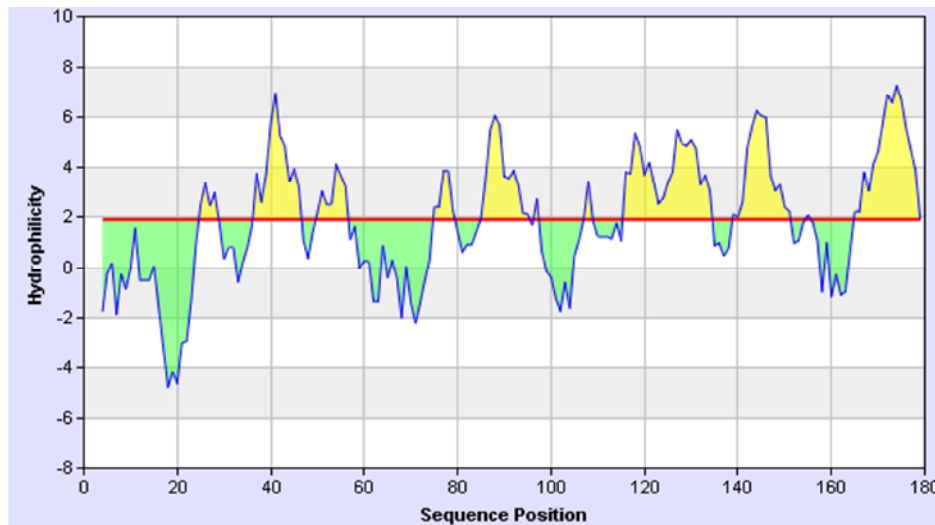

Figure 3 Parker hydrophilicity plot In Parker hydrophilicity prediction scale X axis represents the hydrophilicity score whereas the Y axis represents the sequence position. The average score of hydrophilicity is 1.887 and thus the peptide regions having hydrophilicity score above the average score are hydrophilic and are likely to be present on the surface of the Bnm1 protein. Nine potential hydrophilic regions predicted here are highlighted as yellow color region whereas the green shaded regions could not satisfy the minimum hydrophilicity score and thus are likely to be present within the core of Bnm1 protein.

Hydrophilic regions of the protein are likely to be exposed to outer surface and are most likely to evoke B cell response. To predict the hydrophilic region of the Bnm1, we assigned Parker hydrophilicity prediction [26] method and the predicted hydrophilicity plot shows the hydrophilic regions of Bnm1 protein shaded in yellow color (Figure 3).

Since the single-scale amino acid propensity profiles cannot always predict B cell epitope location reliably as even the best Antigen Prediction method could yield marginally better score than the ROC (receiver operating characteristics) plot [38] so we also performed a machine learning process, BepiPred to predict the antigenic sites that can increase the prediction success which is reviewed elsewhere [39, 40].

3.3 potential B cell epitopes overlap the antigenic sites of Bnm1

Peptides exposed to the outer surface of the protein do not mean that they will react with B cell. To predict the B cell epitopes we employed BepiPred tool which is a antibody epitope prediction program that combines both hidden Markov model and antigenic propensity scale method making the prediction more reliable [27]. BepiPred predicted seven potential B cell epitopes highlighted as yellow marked region for Bnm1 protein sequences (Figure 4) and the maximum score predicted is 2.034. The predicted epitopes fit exactly in the hydrophilic regions predicted from Parker hydrophilicity plot and thus are likely to be exposed on the surface of the protein. A detail of individual B cell epitopes is summarized in Table 3.
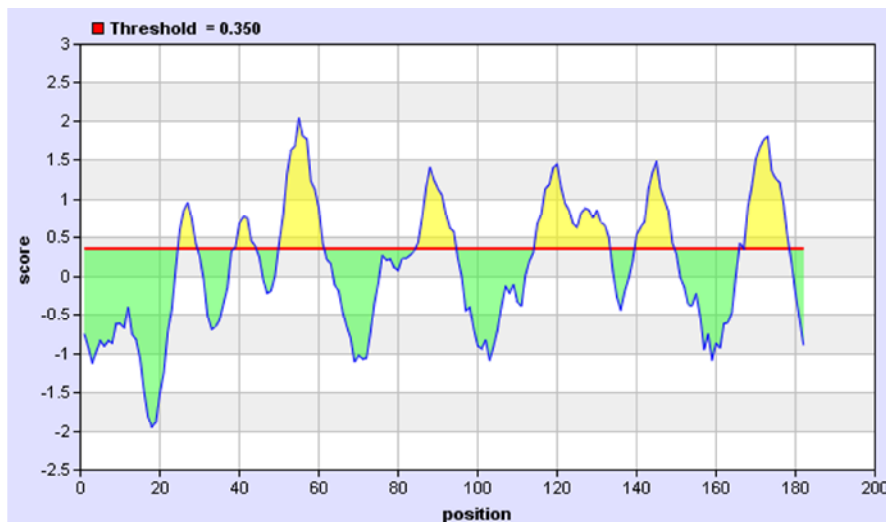

Figure 4 Potential B cell epitopes predicted from BepiPred tool The amino acid sequences having a score above the threshold (0.350) are predicted to be potential B cell epitopes and are highlighted as yellow color. The maximum score predicted here is 2.034

Table 3 Predicted B cell epitope sequences and their position along with their length.

Number	Start position	End position	Epitope sequence	Length of epitope
1	25	29	SASPH	5
2	39	44	RSHDQD	6
3	50	61	LTTNPPTAAPIG	12
4	85	94	ETGKADQTFT	10
5	115	133	KQSPDTAHYDVRSSTDQMK	19
6	140	149	ASKNDQASTT	10
7	166	178	SAADAVDDDDENI	13

3.4 Mapping of the B cell epitopes in the modeled structure confirms their presence on the surface of Bnm1

The predicted 3-D structure of Bnm1(Figure 5) was visualized using Pymol molecular visualization system [41]. Ramachandran plot generated to validate the predicted structure (Figure 5) shows the amino acid distribution in different regions of the plot. A detail
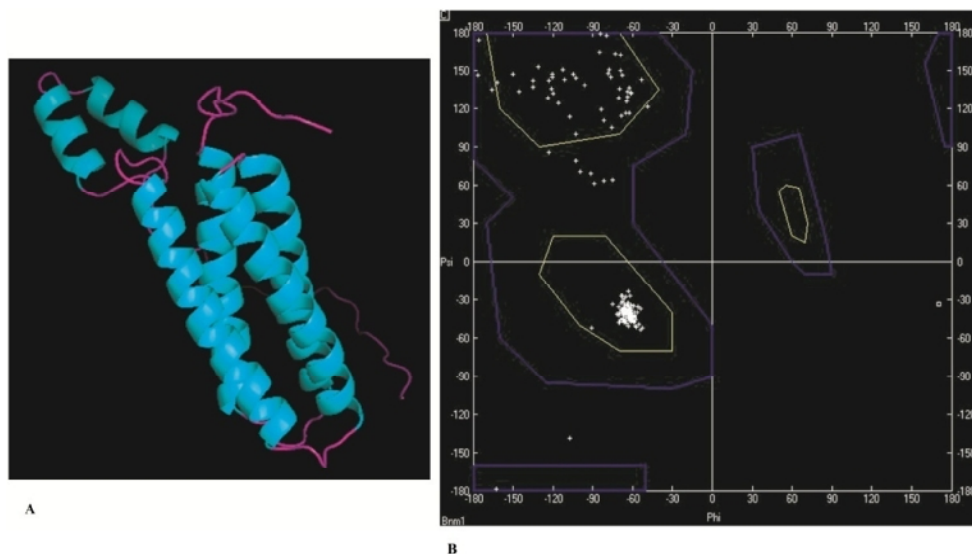
Figure 5 3D structure and its validation using Ramachandran plot for Bnm1 protein (A) Cartoon representation of the predicted structure of Bnm1 protein. This image has been developed using pymol molecular visualization system (B) amino acid residues are distributed in Ramachandran plot. All the amino acid residues lie in the allowed residues in distribution plot making the prediction reliable.

of the distribution of amino acids in different regions is summarized in Table 4. Once the structure was found to be reliable, this 3D structure was chosen as the template to map the

Table 4 Distribution of different amino acid residues in Ramachandran plot

Region of distribution of amino acid residues	Percentage of distribution
Most favored regions	92.9
Additional allowed regions	7.1
Generously allowed regions	0.0
Disallowed regions	0.0

predicted B cell epitopes in Bnm1 protein (Figure 6). The yellow balls (Figure 6) represent the predicted B cell epitopes and all of these yellow colored regions are on the surface of the protein. So what we observed from the primary structure of Bnm1 is consistent in its 3D structure.
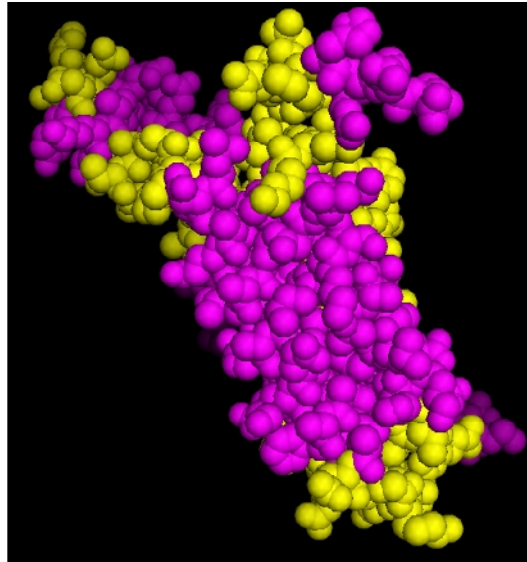
Figure 6 Mapping of B cell epitopes on 3D structure of Bnm1 All the B cell epitopes predicted from BepiPred prediction tool are mapped on the surface of the Bnm1 protein structure where B cell epitopes are highlighted as yellow color and on the other hand the rest of the portion of Bnm1 protein is highlighted as pink color.

3.5 Predicted potential T cell epitopes of Bnm1 shows multiple interactions with DRB1* 0101

We also checked whether Bnm1 can induce T cell response since some allergens can induce both humoral and cell mediated immune response to exert immune-inflammation [42-44]. MHC class II epitope prediction was performed for the highly abundant selected alleles e.g. DRB1*0101, DRB1*0301, DRB1*0401, DRB1*0701, DRB1*1101, DRB1*1301 and DRB1*1501 [31]. The prediction tool retrieved 29 potential candidate T cell epitopes for HLA-DRB1*01:01 allele but only 4 for HLA-DRB1*03:01 allele (Supplementary file 1). There is no allele interaction predicted for DRB1*04:01, DRB1*07:01, DRB1*11:01, DRB1*13:01 and DRB1*15:01 within chosen IC50 range (IC50<25.0). IgE production in response to particular allergens is associated with certain MHC class II alleles, suggesting that particular MHC-peptide interactions might evoke a strong TH2 response and several ragweed pollen allergens induce IgE responses which is associated with haplotypes having HLA-II DRB1*1501 allele [45]. Our result suggests that Bnm1 protein may induce production of IgE response which is largely associated with HLA-DRB1*01:01 allele.

4. Conclusion

To design an effective immune therapy against the allergic response to oilseed rape pollens it is highly needed to know about the array of allergens in its pollen. Our computational approaches strongly suggest that Bnm1, the pollen specific protein, is likely to be an allergen. However, In vitro analyses are warranted to validate the allergenicity of Bnm1.

Acknowledgement

References

1. Lehrer, S.B. and J.E. Salvaggio. *Allergens: Standardization and Impact of BiotechnologyA Review*. in *Allergy and Asthma Proceedings*. 1990: OceanSide Publications, Inc.
2. Vojdani, A., *Detection of IgE, IgG, IgA and IgM antibodies against raw and processed food antigens.* Nutr Metab (Lond), 2009. **6**: p. 22.
3. Woodfolk, J.A., *T-cell responses to allergens.* Journal of allergy and clinical immunology, 2007. **119**(2): p. 280-294.
4. Ozdemir, C., M. Akdis, and C.A. Akdis, *T-cell response to allergens.* 2010.
5. Taketomi, E.A., et al., *Allergens: sources, exposure and sensitization levels, diagnostic tools and immunotherapeutical applications.* 2010.
6. Bufe, A., *The biological function of allergens: relevant for the induction of allergic diseases?* International Archives of Allergy and Immunology, 1998. **117**(4): p. 215-219.
7. Wüthrich, B., *Epidemiology of the allergic diseases: are they really on the increase?* International Archives of Allergy and Immunology, 1989. **90**(1): p. 3-10.
8. Bousquet, J. and P. Burney, *Evidence for an increase in atopic disease and possible causes.* Clinical and experimental allergy: journal of the British Society for Allergy and Clinical Immunology, 1993. **23**(6): p. 484.
9. de Weger, L.A., et al., *Difference in symptom severity between early and late grass pollen season in patients with seasonal allergic rhinitis.* Clinical and Translational Allergy, 2011. **1**(1): p. 1-11.
10. D'amato, G., et al., *Pollen-related allergy in Europe*.* Allergy, 2007. **53**(6): p. 567-578.
11. Platts-Mills, T.A.E. and J.A. Woodfolk, *Allergens and their role in the allergic immune response.* Immunological reviews, 2011. **242**(1): p. 51-68.
12. *Food and Agriculture Organisation of the United Nations : Faostat. .*
13. James, C., *Global Status of Commercialized Biotech/GM Crops: 2010 ISAAA Brief No. 42. ISAAA: Ithaca, NY.* 2011, ISBN 978-1-892456-49-4,[Online], Available: http://www. isaaa. org/resources/publications/briefs/42/executivesummary/default. asp [07.03. 2012].
14. *International Service for the Acquisition of Agri-Biotech Applications: GM Approval Database.*
15. Trinidade, A., et al., *The prevalence of oilseed rape hypersensitivity in a mixed cereal farming population.* Clinical Otolaryngology, 2010. **35**(1): p. 13-17.
16. Chardin, H., et al., *Characterization of high-molecular-mass allergens in oilseed rape pollen.* International Archives of Allergy and Immunology, 2001. **125**(2): p. 128-134.
17. Chardin, H., et al., *Polygalacturonase (pectinase), a new oilseed rape allergen.* Allergy, 2003. **58**(5): p. 407-411.
18. Puumalainen, T.J., et al., *Napins, 2S albumins, are major allergens in oilseed rape and turnip rape.* Journal of allergy and clinical immunology, 2006. **117**(2): p. 426-432.
19. Focke, M., et al., *Identification of Oilseed Rape< i>(Brassica napus)</i> Pollen Profilin as a Cross-Reactive Allergen.* International Archives of Allergy and Immunology, 2003. **132**(2): p. 116-123.
20. de Weerd, N.A., P.L. Bhalla, and M.B. Singh, *Aeroallergens and pollinosis: Molecular and immunological characteristics of cloned pollen allergens.* Aerobiologia, 2002. **18**(2): p. 87-106.
21. Md. Rezaul Islam, M.S.S., Aubhishek Zaman, *A computational assay to design an epitope-based peptide vaccine against chikungunya virus.* Future Virology, 2012. **7**(10): p. 1029-1042.
22. Tchernitchko, D., M. Goossens, and H. Wajcman, *In silico prediction of the deleterious effect of a mutation: proceed with caution in clinical genetics.* Clinical chemistry, 2004. **50**(11): p. 1974-1978.
23. Treacy, B.K., et al., *Bnm1, a Brassica pollen-specific gene.* Plant molecular biology, 1997. **34**(4): p. 603-611.
24. Gasteiger, E., et al., *Protein identification and analysis tools on the ExPASy server.* The proteomics protocols handbook, 2005: p. 571-607.
25. Kolaskar, A. and P.C. Tongaonkar, *A semi-empirical method for prediction of antigenic determinants on protein antigens.* FEBS letters, 1990. **276**(1): p. 172-174.
26. Parker, J., D. Guo, and R. Hodges, *New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites.* Biochemistry, 1986. **25**(19): p. 5425-5432.
27. Larsen, J.E.P., O. Lund, and M. Nielsen, *Improved method for predicting linear B-cell epitopes.* Immunome research, 2006. **2**(1): p. 2.
28. Källberg, M., et al., *Template-based protein structure modeling using the RaptorX web server.* Nature Protocols, 2012. **7**(8): p. 1511-1522.

29. Johansson, M.U., et al., *Defining and searching for structural motifs using DeepView/Swiss-PdbViewer.* BMC bioinformatics, 2012. **13**(1): p. 173.
30. Laskowski, R.A., et al., *PROCHECK: a program to check the stereochemical quality of protein structures.* Journal of applied crystallography, 1993. **26**(2): p. 283-291.
31. Gowthaman, U. and J.N. Agrewala, *In silico tools for predicting peptides binding to HLA-class II molecules: more confusion than conclusion.* Journal of proteome research, 2007. **7**(01): p. 154-163.
32. Nielsen, M., et al., *Quantitative predictions of peptide binding to any HLA-DR molecule of known sequence: NetMHCIIpan.* PLoS computational biology, 2008. **4**(7): p. e1000107.
33. De Groot, A.S., et al., *Activation of natural regulatory T cells by IgG Fc–derived peptide "Tregitopes".* Blood, 2008. **112**(8): p. 3303-3311.
34. Singh, S., et al., *Physical Properties of Intact Proteins May Predict Allergenicity or Lack Thereof.* PloS one, 2009. **4**(7): p. e6273.
35. Bachmair, A., *In vivo half-life of a protein is a function of its.* Science, 1986. **3018930**(179): p. 234.
36. Varshavsky, A., *The N-end rule pathway of protein degradation.* Genes to Cells, 1997. **2**(1): p. 13-28.
37. Hopp, T.P. and K.R. Woods, *Prediction of protein antigenic determinants from amino acid sequences.* Proceedings of the National Academy of Sciences, 1981. **78**(6): p. 3824-3828.
38. Blythe, M.J. and D.R. Flower, *Benchmarking B cell epitope prediction: underperformance of existing methods.* Protein Science, 2009. **14**(1): p. 246-248.
39. Chen, J., et al., *Prediction of linear B-cell epitopes using amino acid pair antigenicity scale.* Amino Acids, 2007. **33**(3): p. 423-428.
40. Söllner, J. and B. Mayer, *Machine learning approaches for prediction of linear B-cell epitopes on proteins.* Journal of Molecular Recognition, 2006. **19**(3): p. 200-208.
41. DeLano, W.L., *The PyMOL molecular graphics system.* 2002.
42. Motta, A., et al., *Phleum pratense pollen starch granules induce humoral and cell-mediated immune responses in a rat model of allergy.* Clinical & Experimental Allergy, 2004. **34**(2): p. 310-314.
43. Lee, S.C., et al., *Regulation of pulmonary T cell responses to inhaled antigen: role in Th1-and Th2-mediated inflammation.* The Journal of Immunology, 1999. **162**(11): p. 6867-6879.
44. Boulet, L.P., *Allergen-derived T cell peptides and late asthmatic responses.* American journal of respiratory and critical care medicine, 2004. **169**(1): p. 2-3.
45. Janeway CA Jr, T.P., Walport M, et al., *Immunobiology: The Immune System in Health and Disease.* . 2001.

Supplementary Files:

Supplementary Table 1 Predicted T cell candidate epitopes having IC50 < 25.0 score

Allele name	Start Position	End position	Core sequence	Peptide sequence	IC50 score
HLA-DRB1*01:01	19	27	LLLVPASAS	TLQLLLVPASASPHM	4.8
HLA-DRB1*01:01	19	27	LLLVPASAS	LQLLLVPASASPHMK	5.0
HLA-DRB1*01:01	17	25	LQLLLVPAS	ITLQLLLVPASASPH	5.4
HLA-DRB1*01:01	19	27	LLLVPASAS	QLLLVPASASPHMKY	6.4
HLA-DRB1*01:01	17	25	LQLLLVPAS	AITLQLLLVPASASP	6.5
HLA-DRB1*01:01	17	25	LQLLLVPAS	AAITLQLLLVPASAS	7.4
HLA-DRB1*03:01	103	111	VVADLKSAN	YLAVVADLKSANLKL	7.7

HLA-DRB1*01:01	160	168	LLDLAASAA	MEKLLDLAASAADAV	9.2
HLA-DRB1*01:01	6	14	VLSTFAAAA	TFSVLSTFAAAAITL	10.0
HLA-DRB1*03:01	103	111	VVADLKSAN	AYLAVVADLKSANLK	10.0
HLA-DRB1*01:01	4	12	FSVLSTFAA	MATFSVLSTFAAAAI	10.3
HLA-DRB1*01:01	6	14	VLSTFAAAA	ATFSVLSTFAAAAIT	10.9
HLA-DRB1*01:01	6	14	VLSTFAAAA	FSVLSTFAAAAITLQ	12.4
HLA-DRB1*01:01	6	14	VLSTFAAAA	SVLSTFAAAAITLQL	12.5
HLA-DRB1*01:01	19	27	LLLVPASAS	LLLVPASASPHMKYI	12.8
HLA-DRB1*03:01	103	111	VVADLKSAN	LAVVADLKSANLKLK	13.0
HLA-DRB1*01:01	10	18	FAAAAITLQ	VLSTFAAAAITLQLL	13.2
HLA-DRB1*01:01	160	168	LLDLAASAA	QMEKLLDLAASAADA	13.3
HLA-DRB1*01:01	160	168	LLDLAASAA	EKLLDLAASAADAVD	13.3
HLA-DRB1*01:01	10	18	FAAAAITLQ	LSTFAAAAITLQLLL	13.4
HLA-DRB1*01:01	71	79	LTIAHAEKT	VMALTIAHAEKTAAF	13.7

Supplementary Table 1 (Continued) Predicted T cell candidate epitopes having IC50 < 25.0 score

Allele name	Start Position	End position	Core sequence	Peptide sequence	IC50 score
HLA-DRB1*01:01	104	112	VADLKSANL	LAVVADLKSANLKLK	14.2
HLA-DRB1*01:01	17	25	LQLLLVPAS	AAAITLQLLLVPASA	14.3
HLA-DRB1*01:01	100	108	YLAVVADLK	HKAYLAVVADLKSAN	14.7
HLA-DRB1*01:01	104	112	VADLKSANL	YLAVVADLKSANLKL	14.7
HLA-DRB1*01:01	100	108	YLAVVADLK	YHKAYLAVVADLKSA	16.9
HLA-DRB1*01:01	100	108	YLAVVADLK	KAYLAVVADLKSANL	17.0

HLA-DRB1*01:01	160	168	LLDLAASAA	KLLDLAASAADAVDD	18.3
HLA-DRB1*03:01	103	111	VVADLKSAN	KAYLAVVADLKSANL	19.8
HLA-DRB1*01:01	71	79	LTIAHAEKT	EVMALTIAHAEKTAA	20.5
HLA-DRB1*01:01	100	108	YLAVVADLK	AYLAVVADLKSANLK	20.7
HLA-DRB1*01:01	10	18	FAAAAITLQ	STFAAAAITLQLLLV	21.8
HLA-DRB1*01:01	71	79	LTIAHAEKT	MALTIAHAEKTAAFV	23.9