American Journal of Infectious Diseases 5 (2): 142-147, 2009
ISSN 1553-6203

Target Identification in Ory S1 Pollen Protein Allergen from *Oryza sativa* in the Course of Construction of Hypoallergenic Vaccines

Ruchi Sharma and A.K. Singh
Department of Botany, Udaya Pratap College, Varanasi,
Veer Bahadur Singh Purvanchal University, Jaunpur, Uttar Pradesh, India

Abstract: Problem statement: Recombinant-based approaches are mostly focused on genetic modification of allergens to produce molecules with reduced allergenic activity and conserved antigenicity, such as hypoallergens. Recombinant allergens represent promising tools for diagnosis and therapy of type I allergy. This approach was probably feasible with every allergen with known amino acid sequence. **Approach:** The primary aim of this study was to determine the consensus epitope from twenty homologous protein sequences of Ory S1 allergenic protein sequence from *Oryza sativa* (indica group) pollen. Molecular modeling calculations had been used to investigate the allergenic protein models for the epitope. **Results:** *Oryza sativa* (japonica), *Phleum pratense*, *Poa pratensis*, *Holcus lanatus*, *Lolium perenne*, *Triticum aestivum*, *Dactylis glomerata* and *Zea mays* were found more closely related (alignment score 1145-812) among all the homologs and investigated further. The major binding pocket comprised an area of 604.5 $Å^2$ and 970 $Å^3$ volume and another key binding pocket had 425.6 $Å^2$ area and 658.8 $Å^3$ volume. The residues found in the key site included ile2, lys13, cys14, ser15, lys16, pro17, ala25, leu26, ile27, tyr40, his41, phe42, asp43, leu44, ser45, gly46, leu47, ala48, met49, ala50, asp55, leu58, arg59, ala61, gly62, ile63, ile64, asp65, gln67, phe68; corresponding to the allergen binding site and the IgE binding epitope given in the title. **Conclusion:** These are the functional sites on the allergenic proteins that can be mutated to develop hypoallergenic vaccine. These sites can be rationalized on the basis of simple arguments that lead to vaccine development, by predicting the structure of the allergenic epitopes and comparative analysis.

Key words: Homology modeling, allergen binding epitope, allergenic proteins

INTRODUCTION

An allergen simply means a harmful immune response elicited by an antigen that is not itself intrinsically harmful. Grass pollens are well known among the health hazardous bio aerosols causing respiratory allergy. Being an important member of the grass family, the rice plants contribute a huge pollen load in agricultural fields during flowering.

Oryza sativa is the cultivated rice, used as staple food by majority of world's population. Pollen allergens of *Oryza sativa* is recognized by the International Union of Immunological Societies (IUIS) official list of allergens[1] which include Ory S1, Ory S7 and Ory S12. Protein Ory S1 has been validated as an allergen on the basis of its recognition by IgE antibodies from allergic individuals[2].

Allergenic site identification can be explained as the residues found in the binding pocket and in IgE binding epitope. Most allergens contain multiple motifs though all the motifs might not prove to be good targets. The target motif must be selective in terms of IgE binding epitopes[3]. The knowledge of molecular nature of allergen-antibody interactions is important to understand the mechanism of conventional immunotherapy, as well as to design alternative immunotherapeutic strategies[4]. The allergy process has been widely studied in last few decades, enhancing the better understanding of allergic problem that affects varied age group.

Vaccination is the most effective technique suggested nowadays for allergy prevention. The molecules developed for vaccination against allergy possess significantly reduced allergenicity in terms of IgE binding and therefore will not lead to anaphylactic reactions upon injection. This approach is probably feasible with every allergen with known amino acid sequence; irrespective of the source (pollen, food, mites) from which it may be derived[5]. Also, the products of agricultural biotechnology should be subjected to a careful and complete safety assessment

Corresponding Author: Ruchi Sharma, Department of Botany, Udaya Pratap College, Varanasi, Uttar Pradesh, India
Tel: 09633137687

for its allergenicity before commercialization. The identification and validation of protein allergens have become more important nowadays as more and more transgenic proteins are introduced into our food chains.

We need to look for the Immunoglobulin Epsilon (IgE) epitopes to confirm the allergy response of any allergenic or transgenic protein. If the bioinformatics methods are standardized and optimized, it may lead to complete exploitation of the transgenic food and for the identification of targets to create hypoallergenic vaccines. Moreover, many attempts have been well documented to predict allergenicity of a query protein by its amino acid sequences[6].

The present study was intended to obtain homologous sequences for Ory S1 allergenic proteins, analyze the homologous protein sequences for allergenic sites in sequence and to validate the obtained targets, pertaining to effective identification of consensus epitope and distinct sequence features.

MATERIALS AND METHODS

Sequence retrieval: Ory S1 Protein (query protein) sequence from *Oryza sativa* was retrieved from NCBI database[7]. Basic local alignment search tool was used to retrieve the homologous sequences by querying against non-redundant database (nr).

Pair wise sequence alignment: Dynamic programming algorithm was used to align individual homologous sequence(s) with the query protein sequence to find the proximity. LALIGN program which implements Huang and Miller algorithm was utilized for this study[8].

Multiple sequence alignment: Nine close homologous sequences were selected for further study based on pairwise alignment score. ClustalW, a neighbor joining algorithm based tool was used to find the Multiple Sequence Alignment (MSA)[9] and a consensus sequence was prepared based on the MSA.

Domain and motif detection: Domain positions of the sequences were identified using InterPro, an integrated resource for protein families, domains and sites that combines a number of databases (referred to as member databases) using diverse methodologies and a varying degree of biological information on well-characterized proteins to derive domain positions in the protein sequence[10]. Multiple Em for Motif Elicitation (MEME) was also used to find common motifs found in the MSA[11].

Protein homology modeling: All 20 allergen homolog sequences and the consensus domain sequences were modeled using homology modeling method using same template (1n10A). MODELLER, a homology modeling tool which implements comparative protein structure modeling by spatial restraints, was utilized for the present study to construct protein models[12,13].

Binding pocket exploration: CASTp server, which used weighted Delaunay triangulation and the alpha complex by shape measurement of the domain structure was used for the present study to locate the binding pockets[14]. This tool measures analytically the area and volume of each pocket and cavity, both in solvent accessible surface (SA, Richards' surface) and molecular surface (MS, Connolly's surface). The obtained pockets were validated based on their functional significance and its contribution to the essentiality and allergenicity of the organisms. Finally, the target's functionality significance and its role in IgE interactions were validated in the light of literature search.

RESULTS

BLAST search was performed to fetch significant homologous sequences of Ory S1 and were short listed for comparison (Table 1). Pairwise alignment of query sequence (Ory S1) with the individual short listed homologs was performed and the results were tabulated (Table 2). The three Dimensional structures were modeled for the homologous sequences and were validated for plausibility. The RMSD value for all the structures with that of the template (1n10A) was calculated to elucidate 3-Dimensional homology (Table 3). Based on RMSD score and similarity score, the most homologous sequences with good modeled structure were selected for target identification. The conserved residues in the domains of the structures were identified (Table 4) and the property of the conserved domains were analyzed (Fig. 1). Based on position-specific probability matrix (Fig. 1) the probabilities of each possible amino acid letter appearing at each possible position in the conserved domain were elucidated. From all these sequences selected, subsets of highly conserved residues were retrieved and a multiple sequence alignment was performed to get the consensus sequences (Fig. 2). Furthermore, the binding analysis were also done all the structures and the potential pocket was identified based on Castp rating (Fig. 3).Moreover, the residues spanning the potential pocket were tabulated in Table 5.

Table1: Sequences and data obtained from NCBI server

Sequence No.	Accession No.	Protein name	Organism	Sequence length
1	AAA86533.1	Ory S1 (Query sequence)	*Oryza sativa*	263
2	NP_001048686.1	Pollen allergen	*Oryza sativa* (japonica)	264
3	CAA81613.1	Pollen allergen Phl pI	*Phleum pratense*	263
4	CAA10520.1	Group I pollen allergen	*Poa pratensis*	263
5	CAA10140.1	Major group I allergen Hol l 1	*Holcus lanatus*	263
6	CAB63699.1	Pollen allergen	*Lolium perenne*	263
7	AAP96760.1	Group 1 allergen Dac g 1.02 precursor	*Dactylis glomerata*	264
8	AAS48882.1	Expansin EXPB4	*Triticum aestivum*	270
9	ABF81662.1	EXPB10	*Zea mays*	269
10	CAC40805.1	Beta expansin B2	*Festuca pratensis*	269
11	ABB83474.*1*	Beta-expansin precursor	*Solanum lycopersicum*	275
12	AAZ08315.1	Putative beta-expansin	*Eucalyptus globulus*	210
13	NP_190182.2	ATEXLA3	*Arabidopsis thaliana*	215
14	ABK93417.1	Unknown	*Populus trichocarpa*	259
15	AAV85475.1	Expansin	*Populus tomentosa*	258
16	ACB45301.1	Beta-expansin EXPB4	*Hordeum vulgare*	273
17	ABJ90221.1	Expansin 2	*Malus hupehensis*	253
18	AAT11859.2	Expansin 1	*Mangifera indica*	260
19	BAC67192.1	Expansin	*Pyrus communis*	253
20	BAC66787.1	Expansin	*Prunus persica*	260

Table 2: Alignment scores for the homologous sequences with the query sequence

Sr. No.	Name of sequence with which query sequence Ory S1 (*Oryza sativa*) is aligned	Identities	Positives	Gaps	Alignment score
1	*Oryza sativa* (japonica cultivar-group)	89	91	0	1145
2	*Phleum pratense* [Phl pI]	70	82	3	910
3	*Poa pratensis* (group I pollen allergen)	67	80	3	881
4	*Holcus lanatus* (Hol l 1)	66	79	3	864
5	*Lolium perenne* (pollen allergen)	67	78	3	860
6	*Dactylis glomerata* (Dac g 1.02 precursor)	69	80	3	889
7	Timothy grass pollen Allergen (Phl P 1)	69	80	3	884
8	*Triticum aestivum* (expansin EXPB4)	64	79	6	830
9	*Zea mays* (EXPB10)	65	77	5	812
10	*Festuca pratensis* (beta expansin B2)	49	66	1	578
11	*Solanum lycopersicum* (Beta-expansin precursor)	42	62	5	442
12	*Eucalyptus globulus* (putative beta-expansin)	41	61	6	420
13	*Arabidopsis thaliana* (ATEXLA3)	35	53	8	261
14	*Populus trichocarpa* (unknown)	30	49	1	246
15	*Populus tomentosa* (expansin)	31	48	14	197
16	*Hordeum vulgare* (EXPB4)	47	65	5	226
17	*Malus hupehensis* (expansin 2)	29	45	15	180
18	*Mangifera indica* (expansin 1)	30	47	13	184
19	*Pyrus communis* (expansin)	29	46	13	176
20	*Prunus persica* (expansin)	30	45	13	179

Table 3: RMSD values for all the structures modeled with 1N10A as template structure

S. No.	Sequence ID	RMSD values (A)
1	AAA86533.1	0.47
2	NP_001048686.1	0.42
3	CAA81613.1	0.38
4	CAA10520.1	0.30
5	CAA10140.1	0.41
6	CAB63699.1	0.30
7	AAP96760.1	0.33
8	AAS48882.1	0.33
9	ABF81662.1	0.42
10	CAC40805.1	0.59
11	ABB83474.1	0.46
12	AAZ08315.1	0.60
13	NP_190182.2	1.05
14	ABK93417.1	0.81
15	AAV85475.1	0.99
16	ACB45301.1	0.49
17	ABJ90221.1	0.91
18	AAT11859.2	1.01
19	BAC67192.1	1.11
20	BAC66787.1	1.08

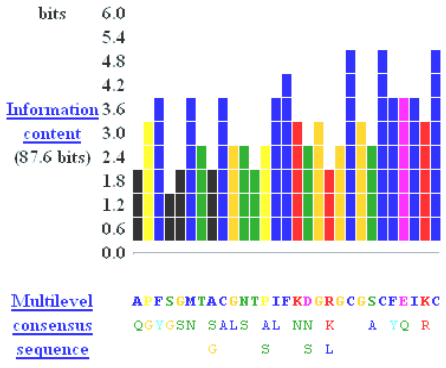

Fig. 1: The information content diagram showing most highly conserved positions in the motif

```
CAB63699.1        ------MASSSS-VLLVVALFAVFLGSAH-GIAKVPPGPNITAE---YGDKWLDAKSTWY  49
AAP96760.1        ------MASSSSVLLVVALFAVFLGSAH-GIPKVPPGPNITAT---YGDKWLDAKSTWY  50
CAA10520.1        ------MASSSS-VLLVVALFAVFLGTAH-GIAKVPPGPNITAT---YGDKWLDAKSTWY  49
CAA10140.1        ------MASSSL-VLLVVALFAVFLGTAH-GIAKVPPGPNITAT---YGDKWLDAKSTWY  49
CAA81613.1        ------MASSSS-VLLVVALFAVFLGSAH-GIPKVPPGPNITAT---YGDKWLDAKSTWY  49
AAS48882.1        ------MASSSSSVLLVAAVLAAVVCGAH-GIAKVPPGPNITASPTSYGNKWLDAKTTWY  53
ABF81662.1        MTVVSIMWSLVQVQVLVAVALSFLVGGAWCGPPKVPPGKNITAN---YGSDWLDAKATWY  57
AAA86533.1        ------MASSSL--LLACVVVAAMVSPSPAGHPKVPPGPNITTS---YGDKWLEARPPGM  49
NP_001048686.1|   ------MASSSL--LLACVVVAAMVSAVSCGPPKVPPGPNITTS---YGDKWLEAKATWY  49

CAB63699.1        GKPTGAGPKDNGGACGYKDVDKAPFNGMTGCGNTPIFKDGRGCGSCFEIKCTKPESCSGE 109
AAP96760.1        GKPTGAGPKDNGGACGYKDVDKAPFNGMTGCGNTPIFKDGRGCGSCFEIKCTKPESCSGE 110
CAA10520.1        GKPTGAGPKDNGGACGYKDVDKAPFSGMTGCGNTPIFKSGRGCGSCFEIKCTKPESCSGE 109
CAA10140.1        GKPTGAGPKDNGGACGYKDVDKPPFSGMTGCGNTPIFKSGRGCGSCFEIKCTKPESCSGE 109
CAA81613.1        GKPTAAGPKDNGGACGYKDVDKPPFSGMTGCGNTPIFKSGRGCGSCFEIKCTKPEACSGE 109
AAS48882.1        GKPTGAGPKDNGGACGYKEVDKAPFHGMTSCGNIPIFKDGRGCGSCFELKCTKPEACSGE 113
ABF81662.1        GKPTGAGPDDNGGGCGYKDVNKAPFNSMGACGNVPIFKDGLGCGSCFEIKCDKPAECSGK 117
AAA86533.1        VRPRVLAPKDNGGACGYKDVDKAPFLGMNSCGNDPIFKDGKGCGSCFEIKCSKPEACSDK 109
NP_001048686.1|   GAPKGAGPKDNGGACGYKDVDKAPFLGMNSCGNDPIFKDGKGCGSCFEIKCSKPEACSDK 109

CAB63699.1        AVTVTITDDNEEPIAPYHFDLSGHAFGSMAKKGEEQKLRSAGELELQFRRVKCKYPDGTK 169
AAP96760.1        AVTVHITDDNEEPIAPYHFDLSGHAFGSMAKKGEEQKLRSAGELELQFRRVKCKYPEGTK 170
CAA10520.1        PVLVHITDDNEEPIAAYHFDLSGKAFGAMAKKGEEQKLRSAGELELKFRRVKCEYPEGTK 169
CAA10140.1        PIVVHITDDNEEPIAAYHLDLSGKAFGAMAKKGEEQKLRSAGELELKFRRVKCEYPKGTK 169
CAA81613.1        PVVVHITDDNEEPIAAYHFDLSGIAFGSMAKKGDEQKLRSAGEVEIQFRRVKCKYPEGTK 169
AAS48882.1        PTMVTITDKNEEPIAPYHFDLSGHAFGSMAKKGEEQKLRDAGEVEIKFRRVKCKYPAGTK 173
ABF81662.1        PVVVYITDMNYEPIAAYHFDLAGTAFGAMAKKGEEEKLRKAGIIDMQFRRVKCKY--GSK 175
AAA86533.1        PALIHVTDMNDEPIAAYHFDLSG---LAMAKDGKDEELRKAGIIDTQFRRVKCKYPADTK 166
NP_001048686.1|   PALIHVTDMNDEPIAAYHFDLSG---LAMAKDGKDEELRKAGIIDTQFRRVKCKYPADTK 166

CAB63699.1        PTFHVEKASNPNYLAILVKYVDGDGDVVAVDIKEKGKDKWIELKESWGAVWRIDTPDKLT 229
AAP96760.1        VTFHVEKGSNPNYLALLVKYVDGDGDVVAVDIKEKGKDKWIALKESWGAIWRVDTPDKLT 230
CAA10520.1        VTFHVEKGSNPNYLALLVKYVTGDGDVVAVDIKEKGKDKWIELKESWGSIWRVDTPDKLT 229
CAA10140.1        VTFHVEKGSNPNYLALLVKYVDGDGDVVAVDIKEKGKDKWIELKESWGAVWRVDTPDKLT 229
CAA81613.1        VTFHVEKGSNPNYLALLVKFVAGDGDVVAVDIKEKGKDKWIALKESWGAIWRIDTPEVLK 229
AAS48882.1        VNFHVEKSSNENYLALVIKFLQGDGDVVGVDIKQKGEDKWTELNESWGAVWRIDTPHKLI 233
ABF81662.1        VTFHLEKGCNPNYLALLVKYVDGDGDIVAVDIKEKGSDTYEPLKHSWGAIWRKDSDKPIK 235
AAA86533.1        ITFHIEKASNPNYLALLVKYVAGDGDVVEVEIKEKGSEEWKALKESWGAIWRIDTPKPLK 226
NP_001048686.1|   ITFHIEKASNPNYLALLVKYVAGDGDVVEVEIKEKGSEEWKALKESWGAIWRIDTPKPLK 226

CAB63699.1        GPFTVRYTTEGGTKSEVEDVIPEGWKADTSYSAK---- 263
AAP96760.1        GPFTVRYTTEGGTKSEVEDVIPEGWKADTSYEAK---- 264
CAA10520.1        GPFTVRYTTEGGTKGEAEDVIPEGWKADTAYASK---- 263
CAA10140.1        GPFTVRYTTEGGTKVEAEDVIPEGWKADTAYESK---- 263
CAA81613.1        GPFTVRYTTEGGTKGEAKDVIPEGWKADTAYESK---- 263
AAS48882.1        GPFSVRYTTEGGTKTVVDDVIPKGWKPDTSYEAKGGY- 270
ABF81662.1        GPITVRLTTEGGTKTVYDDVIPTDWKPNTAYTTK---- 269
AAA86533.1        GPFSVRVTTEGARRSSAEDAIPDPGRRQRV-QVNVQAK 263
NP_001048686.1|   GPFSVRVTTEGGEKIIAEDAIPDGWKADSVYKSNVQAK 264
```

Fig. 2: Multiple sequence alignment for the domain position in nine most related sequences based on alignment score and RMSD value

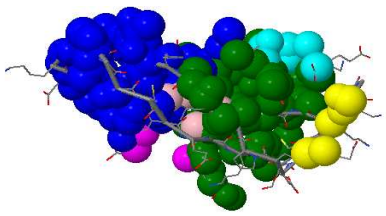

Fig. 3: Binding domains (pockets) using CASTp for domain structure

Table 4: Binding domains positions using PROSITE for nine most similar sequences

S. No.	Sequence	Domain position
1	AAA86533.1	78-154
2	NP_001048686.1	78-154
3	CAA81613.1	78-157
4	CAA10520.1	78-157
5	CAA10140.1	78-157
6	CAB63699.1	78-157
7	AAP96760.1	69-178
8	AAS48882.1	82-161
9	ABF81662.1	86-165

Table 5: Binding domains (pockets) using CASTp for domain structure

Pocket	Area, Å²	Volume, Å³	Color	Amino acid residues (name: position)
6	604.5	970.0	Green	I:12 K:13 C:14 S:15 K:16 P:17 A:25 L:26 I:27 Y:40 H:41 F:42 D:43 L:44 S:45 G:46 L:47 A:48 M:49 A:50 D:55 L:58 R:59 A:61 G:62 I:63 I:64 D:65 Q:67 F:68
5	425.6	658.8	Blue	K:1 D:2 G:3 G:5 C:6 S:8 F:10 I:27 H:28 V:29 T:30 D:31 M:32 N:33 P:36 A:39 Y:40 H:41 F:42 D:43 L:44 S:45
4	29.6	19.9	Cyan	G:46 L:47 D:55 R:59
3	4.7	9.5	Yellow	K:23 K:51 D:52
2	1.2	2.2	Magenta	Y:40 F:68
1	28.4	13.7	Pink	I:12 F:42 L:44

DISCUSSION

The homology search for ORY S1 fetched 20 homologous Sequences ranging from grasses to higher plants (Table 1). The three Dimensional structures were modeled for the homologous sequences and were validated by RMSD value with that of the template

(1n10A) was found between 0.30 and 1.11. Hence, indicative of the plausible model obtained (Table 3). Based on RMSD score (0.30-0.47) and sequence similarity score (1145-812) (Table 2), nine most homologous sequences with good modeled structure were selected: *Oryza sativa* (japonica), *Phleum pratense*, *Poa pratensis*, *Holcus lanatus*, *Lolium perenne*, *Triticum aestivum*, *Dactylis glomerata* and *Zea mays* for target identification. Conserved domains of the structures selected showed identity from position 86-164 (Table 4) in all these sequences which confirms the evolutionary relatedness. These active domains were reported to be of DPBB_1 domain (a conserved region from rare lipoprotein A (RlpA) that has the Double-Psi Beta-Barrel (DPBB) fold. Based on position-specific probability matrix derived (Fig. 1) the probability of each possible amino acid letter appearing at each conserved position was identified and it shows high conservation of cysteine residues and hence confirming the well documented role of disulfide bonds formed between cysteines in IgE binding and in several other allergens[15,16]. Consensus derived from the subset (82-154 residues) (Table 4) of Multiple sequence alignment was subjected to prositescan, which fetch the following domain pattern: K-[DS]-G-[KR]-G-C-G-S-C-F-E-I-K-C-[ST]-K-P-E-[AS]-C-S-[DG]-[EK]-[AP]-[AIV]-x-[IV]-x-[IV]-T-D-x-N-[DE]-E-P-I-A-[AP]-Y-H-[FL]-D-L-S-G-x-A-[FM]-[AG]. The consensus domain was modeled based on homology with (1n10A) as template. Structure for the domain position showed a total of six binding pockets, out of which the first major pocket had an area of 604.5 Å^2 (Fig. 3) and 970 Å^3 volume,indicative of sufficient volume for antigen-antibody interactions. As these are common pocket from nine structures, the outcome of this strategy can be used to identify common allergenic epitopes for similar structures at a single stretch. The role of conserved cysteines (Table 5) shall also play a major role in determining the allergenicity and this falls in line with the previously documented studies[15,16]. Hence, the procured consensus region shall be utilized for effective vaccine design against food allergens.

CONCLUSION

Consensus epitope identification using the available allergenic region has geared up the swiftness to fulfill the demands of the patients. If bioinformatics approaches are standardized and optimized, it can be used for the rapid identification of potential antigenic regions to develop hypoallergenic vaccines. The present study shows that allergenic epitopes have some common amino acid conservations in the allergenic domain positions: 82-154, which are conserved with hydrophobic residues. Amino acids cys, lys, gly, ser and pro was found more conserved in the allergenic motif as well as in binding pocket. Cysteine residue highly conserved at four positions shall play a crucial role in antigenicity. We present the results of this study to the medical community for vaccine development.

REFERENCES

1. Marsh, D.G., L. Goodfriend, T.P. King, H. Lowenstein and T.A. Platts-Mills, 1986. Allergen nomenclature. Bull, 64: 767-774.
2. Russell, S.D., P.L. Bhalla and M.B. Singh, 2008. Transcriptome-based examination of putative pollen allergens of rice (*Oryza sativa* ssp. japonica) Mol. Plant, 1: 751-759. DOI: 10.1093/mp/ssn036
3. Kong, T.S.T., L. Tham and K.W. Choo, 2006. Improved prediction of allergenicity by combination of multiple sequence motifs. Silico Biol., 7: 6-6. http://www.bioinfo.de/isb/2006/07/0006/main.html
4. Smith, A.M. and M.D. Chapman Clin, 1997. Localization of antigenic sites on Der p 2 using oligonucleotide-directed mutagenesis targeted to predicted surface residues. Exp. Allerg., 27: 593-599. DOI: 10.1111/j.1365-2222.1997.tb00750.x
5. Kumar and Shelokar, 2008. An SVM method using evolutionary information for the identification of allergenic proteins. Bioinformation, 2: 253-256. http://www.bioinformation.net/002/005700022008.pdf
6. Ferreira, F., C. Ebner, B. Kramer, G. Casari, P. Briza and A.J. Kungl *et al*., 1998. Modulation of IgE reactivity of allergens by site-directed mutagenesis: Potential use of hypoallergenic variants for immunotherapy. FASEB. J., 12: 231-242. http://www.fasebj.org/cgi/content/abstract/12/2/231
7. National Centre for Biotechnology Information; www.ncbi.nlm.nih.gov
8. Huang and Miller, 1991. The lalign program implements the algorithm of Huang and Miller. Adv. Appl. Math., 12: 337-357. http://www.ch.embnet.org/software/LALIGN_form.html
9. Larkin, M.A., G. Blackshields, N.P. Brown, R. Chenna and P.A. McGettigan *et al*., 2007. Clustal W and Clustal X version 2.0. Bioinformatics, 23: 2947-2948. http://www.ncbi.nlm.nih.gov/pubmed/17846036
10. Quevillon, E., V. Silventoinen, S. Pillai, N. Harte, N. Mulder, R. Apweiler and R. Lopez, 2005. InterProScan: Protein domains identifier. Nucleic Acids Res., 33: W116-W120. http://www.ncbi.nlm.nih.gov/pubmed/15980438

11. Bailey, T.L. and C. Elkan, 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proceedings of the 2nd International Conference on Intelligent Systems for Molecular Biology, Aug. 1994, AAAI Press, Menlo Park, California, pp: 28-36. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.56.5763
12. Sali, A. and T.L. Blundell, 1993. Comparative protein modeling by satisfaction of spatial restraints. J. Mol. Biol., 234: 779-815. http://www.ncbi.nlm.nih.gov/pubmed/8254673
13. Fiser, R.K. Do and A. Sali, 2000. Modeling of loops in protein structures. Protein Sci., 9: 1753-1773. http://www.ncbi.nlm.nih.gov/pubmed/11045621
14. Dundas, J., Z. Ouyang, J. Tseng, A. Binkowski, Y. Turpaz and J. Liang, 2006. CASTp: Computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. Nucleic Acid Res., 34: 116-118. DOI: 10.1093/nar/gkl282
15. Lombardero, M., P.W. Heymann, T.A. Platts-Mills, J.W. Fox and M.D. Chapman, 1990. Conformational stability of B cell epitopes on group I and group II *Dermatophagoides* spp. allergens: Effect of thermal and chemical denaturation on the binding of murine IgG and human IgE antibodies. J. Immunol., 144: 1353-1360. http://www.ncbi.nlm.nih.gov/pubmed/1689351
16. Olsson, S., M. van Hage-Hamsten and P. Whitley, 1998. Contribution of disulphide bonds to antigenicity of Lep d 2, the major allergen of the dust mite *Lepidoglyphus* destructor. Mol. Immunol., 35: 1017-1023. DOI: 10.1016/S0161-5890(98)00101-1
17. Gasteiger E., A. Gattiker, C. Hoogland, I. Ivanyi, R.D. Appel and A. Bairoch, 2003. ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res., 31: 3784-3788. http://www.ncbi.nlm.nih.gov/pubmed/12824418