Original Article

Immunoinformatics and Similarity Analysis of House Dust Mite Tropomyosin

Mohammad Mehdi Ranjbar¹, Sedigheh Nabian², Nayeb Ali Ahmadi³*, Khodayar Ghorban⁴, Alireza Sazmand⁵, Maryam Dadmanesh⁶, Seyed HosseinHekmatimoghaddam⁷

¹Razi vaccine and Sera Institute, Karaj, Iran
²Department of Parasitology, Tick and Tick-Borne Diseases Center, Faculty of Veterinary Medicine, University of Tehran, Tehran, Iran
³Proteomics Research Center, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran
⁴Department of Immunology, AJA University of Medical Sciences, Tehran, Iran
⁵Department of Agriculture, Payame Noor University-Yazd Branche, Yazad, Iran
⁶Department of infectious diseases, AJA University of Medical Sciences, Tehran, Iran
⁷Department of Pathology, Shahid Sadoughi University of Medical Sciences, Yazd, Iran

Abstract

Background: Dermatophagoides farinae and Dermatophagoides pteronyssinus are house dust mites (HDM) that cause severe asthma and allergic symptoms. Tropomyosin protein plays an important role in mentioned immune and allergic reactions to HDMs. Here, tropomyosin protein from Dermatophagoides spp. was comprehensively screened in silico for its allergenicity, antigenicity and similarity/conservation.

Materials and Methods: The amino acid sequences of D. farinae tropomyosin, D. pteronyssinus and other mites were retrieved. We included alignments and evaluated conserved/ variable regions along sequences, constructed their phylogenetic tree and estimated overall mean distances. Then, followed by with prediction of linear B-cell epitope based on different approaches, and besides in-silico evaluation of IgE epitopes allergenicity (by SVMc, IgE epitope, ARPs BLAST, MAST and hybrid method). Finally, comparative analysis of results by different approaches was made.

Results: Alignment results revealed near complete identity between D. farina and D. pteronyssinus members, and also there was close similarity among Dermatophagoides spp. Most of the variations among mites' tropomyosin were approximately located at amino acids 23 to 80, 108 to 120, 142 to 153 and 220 to 230. Topology of tree showed close relationships among mites in tropomyosin protein sequence, although their sequences in D. farina, D. pteronyssinus and Psoroptes ovis are more similar to each other and clustered. Dermanyssus gallinae (AC: Q2WB10) has less relationship to other mites, being located in a separate branch. Hydrophilicity and flexibility plots revealed that many parts of this protein have potential to be hydrophilic and flexible. Surface accessibility represented 7 different epitopes. Beta-turns in this protein are with high probability in the middle part and its two terminals. Kolaskar and Tongaonkar method analysis represented 11 immunogenic epitopes between amino acids 7-16. From comparative analysis of predicted probable consensus epitope regions by machine learning approaches these epitopes were gained: AA23-48, AA59-80, AA91-110, AA114-143, AA154-168, AA182-200, AA208-225, and AA254-272. Prediction of allergenic proteins by AlgPred server showed 10 matches for IgE epitope, and prediction by hybrid approach showed that IgE epitope is undoubtedly the major allergen.

Conclusion: Immunoinformatic approaches in allergenic protein analysis are now reliable tools for explanation/interpretation of clinically observed complexities. Results of present study, would help in HDM immunotherapy against several species of parasites as a wide range epitopic desensitization or prevention (vaccine) regime.
Keywords: House dust mite, Tropomyosin, Allergenicity, Antigenicity, Similarity

*Corresponding Author: Nayeb Ali Ahmadi, Proteomics Research Center, and Department of Medical Laboratory Technology, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran. Email: nayebalia@yahoo.com; nayebalia@sbmu.ac.ir


Introduction

Allergy is a hypersensitivity reaction of the immune system to specific external substances called allergens, which are normally harmless substances in the environment. Environmental factors responsible for allergy include animal dander, foods, house dust mites (HDMs), pollen, insects, and chemical substances, to just name a few. Allergy to HDMs is popular and affects millions of people around the world. This allergy is not usually a direct life threatening danger, but could be a trigger for development of asthma, and may eventually lead to death due to respiratory complications1,3.

HDMs belong to class Arachnida and include Dermatophagoides farina and Dermatophagoides pteronyssinus, which are defined as one of the major sources of aeroallergens that cause type I hypersensitivity disease (which sensitize and induce rhinitis, asthma, or atopic dermatitis) in a large portion of the world population2,4. The HDM extract used in allergy diagnosis and therapy is a complex mixture of allergens and non-allergen components from mite5. More than 30 different proteins from mite can bind human IgE6.

Tropomyosin (TPM) protein (group 10 allergen from HDMs) has actin regulator activity and plays an important role in immune and allergic reactions. Despite its allergenicity; it is a candidate for vaccine production from some animal parasites7,9. This protein belongs to a family of highly conserved proteins with multiple isoforms found in both muscle and non-muscle cells of all species of vertebrates and invertebrates. Besides its role in the contractile activity of these cells, it also helps in regulation of cell morphology and motility. Its native structure consists of two parallel alpha-helical TPM molecules that are wound around each other forming a coiled-coil dimer10,11.

B-cell epitope prediction plays an important role in vaccine design, allergy and immunodiagnostics research, and also in understanding immune system functions. Experimental methods for prediction of epitopes are costly and timely processes to find the antigen-antibody reaction sites. Computational methods accelerate reliable prediction of epitopes for different uses in biology. It is also a critical challenge in immunoinformatics and computational immunology. Such methods are cost effective and less time-consuming. B-cell epitopes can be linear or conformational by their structure. Linear epitopes consist of linear sequence of amino acids that can be recognized by either a particular antibody molecule or a particular B-cell receptor of the immune system. Parameters such as hydrophilicity, flexibility, accessibility, turns, exposed surface, and antigenic propensity of polypeptide chains are physico-chemical properties that have been correlated with the location of continuous epitopes12. Amino acid scale-based methods and machine learning methods are two main strategies that have various approaches and are used for prediction of linear B-cell epitopes9. Prediction and evaluation of allergenicity and antigenicity by hybrid approaches usually yields more accurate and reliable results13, and we just used one such method.

Aims of this study were 1) Obtaining accurate details about allergenicity, antigenicity and epitope mapping of TPM in HDMs by various approaches; 2) Describe physico-chemical properties of this protein in deep insight; 3) Preparing a sound basis for integrated and uniform epitopic desensitization and prevention (vaccine) regimes, relying on cross-reactivity of TPM protein among mites; and 4) evaluation of similarity, conservation and evolutionary relationships in dust mites (not only HDMs) by phylogenetic tree.
Methods

Retrieving alignment and also conserved and variable regions of sequences
Complete protein sequences of *D. farinae* (AC: Q23939), *D. pteronyssinus* (AC: O18416) (two house dust mites), and other mites were retrieved from Uniprot KB database. The obtained sequences were aligned using Clustal X\textsuperscript{14}, analyzed and trimmed in CLC sequence viewer software version 6.6.2. Then, very short sequences and areas with ambiguous alignment or containing poly-N stretches were excluded from the analyses. The most highly conserved and variable regions were evaluated by CLC software, as depicted in figure 1.

Constructing the phylogenetic tree and overall mean distances
Selected, aligned and edited sequences directed according to phylogenetic tree were constructed by MEGA 5.3 software package\textsuperscript{15}, as seen in figure 2. We added homo sapiens tropomyosin alpha-1 chain sequence which was used as the out-group to mites tropomyosin dataset. All positions containing gaps and missing data were eliminated. Trees were constructed using the neighbor-joining (NJ) algorithm under the global gap removal option and Kimura’s two-parameter substitution model\textsuperscript{16}. Robustness of phylogenetic analysis was measured by bootstraps analysis with 10,000 replications. The percentage of replicate trees in which the associated taxa are clustered together in bootstraps test is shown next to the branches\textsuperscript{17}. Then, the number of amino acid substitutions per site for overall means distances was calculated. Analyses were conducted using the Poisson correction model\textsuperscript{18}.

Prediction of linear B-cell epitope based on physico-chemical properties
These physico-chemical properties are shown in figure 3.

Antigenicity prediction based on hydrophobicity
Tropomyosin was scanned for hydrophobicity and Parker hydrophilicity index\textsuperscript{19} with window size being seven. Hydrophobicity (or hydrophilicity) plots are designed to display the distribution of polar and non-polar residues along a protein sequence. Most commonly, this analysis has the goal of predicting membrane-spanning segments (highly hydrophobic) or regions that are likely exposed on the surface of proteins (hydrophilic domains) and therefore found useful for identifying potentially antigenic segments. IEDB analysis resource tools were applied for achieving this goal (http://tools.immuneepitope.org/tools/bcell/).

Antigenicity prediction based on assessment of solvent accessibility regions
The accessibility profile was developed using the formulae mentioned by Emini\textsuperscript{20}. Each part of the sequences with surface probability greater than 1.0 indicates an increased probability for being found on the surface. These data and assessments may be useful for prediction of the peptides participating in antigenic activity, surface region peptides and useful domain(s) in the sequence.

Antigenicity prediction based on flexibility
We also concentrated on the flexibility data of protein to increase the prediction accuracy.
This was done with Karplus and Schulz flexibility prediction\textsuperscript{21}. The calculation based on a flexibility scale is similar to classical calculations, except that the center is the fourth amino acid of the seven amino acid window.

Antigenicity prediction based on secondary structure (beta-turn prediction)
Secondary structures of HDM tropomyosin protein sequence were determined to predict the most probable regions and structures involved in antigenicity. The Chou and Fasman method was applied for prediction of the tropomyosin secondary structure\textsuperscript{22}. Using these information parameters, the likelihood of a given residue assuming each of the four possible conformations alpha, beta, reverse turn, or coil was calculated, and the conformation with the most likelihood was assigned to the residue.

Antigenicity prediction based on Kolaskar and Tongaonkar method
Tropomyosin antigenic epitopes were determined using the method of Kolaskar and Tongaonkar\textsuperscript{23}, as is seen in figure 4. This prediction is based on a semi-empirical approach, depending on physico-chemical properties of amino acid residues (i.e. hydrophilicity, accessibility and flexibility). This approach has a good efficiency by detecting antigenic peptides with about 75% accuracy.
B-cell epitope prediction by machine learning approaches

Recently, several methods for B-cell epitope prediction using machine learning approaches have been published. We used four of them in our study as a hybrid method; they included hidden Markov model, feed forward and recurrent neural network, and subsequence Kernel based support vector machine (SVM), which were used in BepiPred, ABCPred, BCPred and again ABCPred, respectively.

In-silico evaluation of allergenicity

World Health Organization (WHO) and Food and Agriculture Organization (FAO) proposed guidelines to assess the potential allergenicity (http://www.fao.org/es/ESN/food), and in this study we used AlgPred server (http://www.imtech.res.in/raghava/algpred/). AlgPred allows prediction of allergens based on similarity of known epitopes with any region of the target protein. Mapping of IgE epitope(s) feature of the server allows user to locate the position of epitope in their protein, and also allows predicting allergens based on SVM modules using amino acid or dipeptide composition. It facilitates BLAST search against 2890 allergen-representative peptides (ARPs) and assigning a protein allergen if it has a BLAST hit. Finally, the hybrid option of server allows predicting allergen using combined approach (SVMc+IgE epitope+ARPs BLAST+MAST).

Comparative analysis

Finally, we compared all of the analyses mentioned above, for interpretation of antigenicity and allergenicity regions of tropomyosin.

Results

Alignment

Alignment results revealed that there was near
complete identity between D. farina and D. pteronyssinus members, and also among *Dermatophagoides* spp. (Fig. 1). Most of the variations among mites tropomyosin were located at about amino acids 23 to 80, 108 to 120, 142 to 153 and 220 to 230.

**Phylogenetic Tree**

For the phylogenetic analysis, tropomyosin sequences of mites (12 sequences) were aligned, compared and edited using CLC sequence viewer 6.6.2. MEGA 5.3 software package was used for construction of phylogenetic tree and calculation of overall mean distances (Fig. 2). Topology of tree shows the close relationships among mites in tropomyosin sequence, although the sequences in *D. farina*, *D. pteronyssinus* and *Psoroptes ovis* are more similar to each other than to other mites and clustered. The nearest relatives to the mentioned cluster are *Glycyphagus domesticus* and *Lepidoglyphus destructor*. *Dermanyssus gallinae* (AC: Q2WB10) has less relationship to other mites, being located in a separate branch. Our out-group, as is shown, was Homo sapiens tropomyosin alpha-1 chain.

Furthermore, estimates of average evolutionary divergence over sequence pairs in mites showed overall divergence over parasite sequence pairs (d) 0.087 with standard error (SE) estimate being 0.010. This reflects highly conservation of this protein among mites. Pair wise alignments and similarity analysis showed that there are highly conserved areas and significant cross reactivity between *Dermatophagoides* spp. in tropomyosin. So, we used *D. farina* as representative of this genus for further analysis.

**Antigenicity prediction based on different methods**

Hydropilic plot (Fig.3A) revealed that many parts of this protein have potential to be hydrophilic (locations above 3 are supposed to have hydrophilicity potential). Careful examination of the surface accessibility predicted peptides diagram using Emini algorithm (Fig. 3B) represents these epitopes: AA70-76 (KLEEEKE), AA98-104 (DLERSEE), AA119-126 (SADESERM), AA157-162 (DADRKY), AA177-184
Table 1: Predicted peptides by Kolaskar-Tongaonkar algorithm. Arrangement in table from ascending to descending is based on their position in sequence. Flexible length for peptides was allowed.

<table>
<thead>
<tr>
<th>No.</th>
<th>Start position</th>
<th>End Position</th>
<th>Peptide</th>
<th>Peptide Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>43</td>
<td>51</td>
<td>VRALOKKKIQ</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>57</td>
<td>68</td>
<td>LDQVQEQLSAAN</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>82</td>
<td>97</td>
<td>EGDVAALNRRIQLIEE</td>
<td>16</td>
</tr>
<tr>
<td>4</td>
<td>105</td>
<td>116</td>
<td>RLIATAKLEEA</td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td>162</td>
<td>175</td>
<td>YDEVARKLAMVEAD</td>
<td>14</td>
</tr>
<tr>
<td>6</td>
<td>188</td>
<td>194</td>
<td>SKIVELE</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td>196</td>
<td>203</td>
<td>ELRVVGNN</td>
<td>8</td>
</tr>
<tr>
<td>8</td>
<td>206</td>
<td>212</td>
<td>SLEVSEE</td>
<td>7</td>
</tr>
<tr>
<td>9</td>
<td>243</td>
<td>256</td>
<td>ERSVQKLQKEVDRL</td>
<td>14</td>
</tr>
<tr>
<td>10</td>
<td>258</td>
<td>264</td>
<td>DELVHEK</td>
<td>7</td>
</tr>
<tr>
<td>11</td>
<td>267</td>
<td>273</td>
<td>YKSISDE</td>
<td>7</td>
</tr>
</tbody>
</table>

Table 2: Predicted B-cell epitopes by different servers that used machine learning approaches. Threshold or specificity was 85% in all servers (except Bepipred). Fixed length epitope prediction was 16 (except Bepipred and FBCPred). Arrangement from ascending to descending is based on score (except Bepipred).

<table>
<thead>
<tr>
<th>Server name</th>
<th>Epitopes</th>
<th>Positions</th>
<th>Features used</th>
<th>Machine Learning Technique</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABCpred Server</td>
<td>HRSDITEERMDGLENQ, KEARMMAEADDRKYDE, SERMRKMLEHRSITDE, DELVHEKEKYKISDE, IQLEIEDLERSEERLK, RAEIAEQKARDANLRA, SQSADERSERMKMLEH</td>
<td>132-148, 149-165, 123-139, 258-274, 92-108, 21-37, 117-133</td>
<td>Hydrophilicity, accessibility, flexibility, turns, antigenicity, polarity</td>
<td>Feed Forward and recurrent Neural network</td>
</tr>
<tr>
<td>BCPREDs Server 1.0</td>
<td>EVSEEKAQQREEAYEQ, TGSKIVEEEELREVQ, QLIEIEDLERSEERLK, EKSEEVEALQKIQIQ, EEAQSDAEERMkRM, VDRLEDELVKEKEK, MEAIKKKQOAMKLEK</td>
<td>208-224, 185-201, 93-109, 37-53, 114-130, 253-269, 1-17</td>
<td>Hydrophilicity, accessibility, flexibility, turns, antigenicity, Amino Acid Pair(AAP) antigenicity scale</td>
<td>Subsequence kernel based SVM</td>
</tr>
<tr>
<td>AAP (Cheng et al., 2007) (in BCPREDs Server 1.0)</td>
<td>EEKAQSQREEAYEQQR, QVQEQLSAANTKLEEK, EDADRYDEVARKLAM, IEEDLERSEERKIAAT</td>
<td>211-227, 59-75, 156-172, 95-111</td>
<td>Hydrophilicity, accessibility, flexibility, turns, antigenicity, Amino Acid Pair (AAP) antigenicity scale</td>
<td>Support Vector Machine (SVM)</td>
</tr>
</tbody>
</table>
AA263-268 (EKEKYK). Karplus and Shulz flexibility prediction plot (Fig. 3C) showed that this protein is highly flexible in different parts of its sequence. Finally, beta-turn prediction (Fig. 3D) reflects that turns in this protein are in the middle parts and its two terminals with high probability.

Considering the methods mentioned above, perhaps the simplest approach for prediction of antigenic determinants is Kolaskar and Tongaonkar23, which is based on the occurrence of amino acid residues in experimentally determined epitopes. Antigenicity prediction plot of tropomyosin by use of the Kolaskar-Tongaonkar algorithm is represented in Figure 4. Also, details of predicted peptides by Kolaskar-Tongaonkar algorithm are arranged in informative Table 1.

Kolaskar-Tongaonkar method evaluates hydrophilicity, accessibility and flexibility together to predict epitopes.

B-cell epitope prediction by machine learning approaches

Details of prediction by different servers are summarized in table 2. Eventually, comparative analysis of probable
consensus epitope regions by machine learning approaches, these epitopes were obtained: AA23-48, AA59-80, AA91-110, AA114-143, AA154-168, AA182-200, AA208-225, and AA254-272.

Allergenicity analysis
Chosen prediction approaches were mapping of IgE epitopes and PID, SVM module based on amino acid composition, SVM module based on dipeptide composition, BLAST search on allergen representative peptides (ARPs) and hybrid approach (SVMc+IgE epitope+ARPs, BLAST+MAST), respectively. Additionally, prediction by hybrid approach showed that this protein is undoubtedly allergen. Analysis results are shown in table 3 and 4.

Discussion
Various studies have reported that exposure to HDM allergens may be a primary cause or a risk factor for development of asthma. About 50% of allergic patients and up to 80% of asthmatic children are sensitized to mite allergens. Mite allergic airway diseases are conventionally defined as a type-I allergy, mediated by Th2 cells and IgE, that is characterized by chronic airway inflammation.

Immunoinformatic approaches in allergenic protein analysis are now reliable tools which help to provide explanation/interpretation for clinically observed diseases in terms of appearances, cross-reactivity, and co-variation of sensitization. A number of methods have been developed to predict allergens, each having its own merits and demerits. In case of AlgPred, a systematic attempt has been made to integrate various approaches in order to predict allergenic proteins with high accuracy. Results of allergenicity prediction by AlgPred are similar to other reports on allergenicity evaluation of tropomyosin in parasites and non-vertebrates.

Comparative analysis of predicted IgE epitopes with

<table>
<thead>
<tr>
<th>Protein name</th>
<th>IgE epitope</th>
<th>Sequence matched</th>
<th>Positions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TPM</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
predicted consensus probable epitope regions by different machine learning approaches showed that all IgE epitopes are in the range of predicted consensus probable epitope regions by machine learning approaches, except AA23-48 and AA208-225. These two epitopes could be just immunogenic (not allergenic) or maybe allergen epitopes that are not yet known. Another comparative analysis of predicted IgE epitopes with predicted epitopes by Kolaskar and Tongaonkar method reveal that except allergenic epitopes of AA151-164 and AA133-148, other epitopes were predicted by Kolaskar and Tongaonkar. Also, epitope AA43-51 was only seen in Kalkaska and Tongaonkar method results. Sharique et al., 2012 also evaluated tropomyosin similarity and allergenicity. A similar analysis was performed using reported B-cell IgE-binding epitopes from Met e1 (shrimp allergen) and Bla g7 (cockroach allergen) with other invertebrate tropomyosin. The percent identity in linear sequences was higher than 35% in mites, crustaceans, and cockroaches. The polar and hydrophobic regions in these groups were highly conserved. Also, results reflect that tropomyosin has different segments that are responsible for allergenicity and immunogenicity. Linear prediction servers (exactly machine learning approaches) were successful in prediction, and most of the epitopes predicted as B-cell epitope are allergenic.

Sequence analysis based on alignment, evaluation of conserved/variable regions and phylogenetic tree construction revealed that there is high conservation between mite species in this protein. It is in accordance with previous studies that identified tropomyosin as a conserved and cross-reactive allergen (sharing similar IgE binding epitopes) between mites and other invertebrates. Therefore, it may be a cause of clinically reported cross-reactivity.

In addition, tropomyosin is present in mites, ticks and insects. So, results of allergenicity and antigenicity analysis would help in HDM immunotherapy as a wide range epitopic desensitization or prevention regime against several species of parasites, although they might have other hosts. Besides, a deep insight to sequence conservation/variability, similarity and phylogenetic relationships is crucial for study of sequence evolution and for identification of functional regions of the protein.

Our Study revealed some unique and valuable immunoinformatic aspects of tropomyosin in Dermatophagoides genus, but due to lack of the protein sequence data in mites, definition of a robust phylogeny and similarity remained unachieved. Here, various approaches with different features for the protein linear B-cell epitope prediction were studied. By combining various features used in all of these approaches and doing a principal component analysis, several features which do not play a major role in epitope prediction were filtered out. There is a need for more studies about allergens of HDM and its role in allergies and asthma. Immunoinformatic approaches help toward deep understanding of the allergic response to HDMs. Comprehensive and comparative assessment of allergenicity, antigenicity, epitope mapping and phylogenetic relationships has never been easy, especially when we attempt to make statements from different aspects about a protein.

Acknowledgment

We appreciate the Proteomics Research Center head and all their colleagues who cooperated in this work.

References


