Molecular characterization and In silico analysis of Sorghum Panallergens: Profilin and Polcalin

B Chandra Sekhar¹, Ch Sachin¹, BV Raman² & JS Bondili¹*

¹Department of Biotechnology, K L University, Vaddeswaram, Guntur, Andhra Pradesh-522 502, India.
²Department of Basic Sciences, Madanapalle Institute of Technology and Sciences, Madanapalle, Andhra Pradesh-517 325, India.

Received 30 October 2014; Revised 23 March 2015

In India, 20-30% of the human population suffer from allergic rhinitis and 15% of them develop asthma. Plant pollens are one of the causative aeroallergens and are mixture of a number of molecules including major and minor allergens (Panallergens). Profilin and polcalcin are the known pollen specific panallergens. Allergenicity of the Sorghum plant in Andhra Pradesh was found to be 54.9%. But the allergens responsible have not been characterized well. This study highlights identification and molecular characterization of Sorghum bicolor profilin (Sorb PF) and S. bicolor polcalcin (Sorb PC) allergen genes based on homology. The coding sequences of the two genes were PCR amplified from the cDNA constructed from Sorghum pollen total RNA. The gene sequences were deposited in NCBI, KC427126 and KC427125. Recombinantly expressed histidine tag (His-tag) purified Sorghum polcalcin and profilin confirmed 9 and 14 kDa proteins, respectively. Based on multiple sequence alignment and phylogenetic analysis, Sorghum polcalcin and profilin were found to be closely related with Cynodon dactylon, Phleum pratense and Oryza sativa grass species. In silico Algpredbased screening of SorbPF and SorbPC showed an allergenicity score of 1.149 and 0.879, respectively. The structure of two EF-hand sequences (DTNGDGKISLSEL and DTDGDGFIDFNEF) of SorbPC showed an exact match with Phlp7. It is concluded that Sorghum recombinant profilin and polcalcin proteins can be of potential use in developing diagnostic kits for allergenicity to Sorghum pollen grains.

Keywords: Allergens, Epitopes identification, Pollen allergens, SorbPF, SorbPC, Sorghum bicolor.

Aeroallergens, including fungal pathogens (spores), play a main role in the pathogenesis of allergic respiratory diseases, such as bronchial asthma, allergic rhinitis and atopic dermatitis¹². Such allergens are considerably increasing in developing countries like India and all over the world. According to the survey held at 18 different centers in the country, 20-30% of Indian population suffer from allergic rhinitis and 15% of them develop asthma¹. Pollen grain proteins are identified as important aeroallergens and are classified into 29 protein families and most of them belong to the expansin, profilin and polcalcin³. The role of different pollen allergens vary with pollution, climatic factors, environmental conditions and degree of exposure. The variance of airborne pollen allergens exists in accordance with different climatic conditions, variations in the diurnal and seasonal conditions. Monitoring these allergens in any important geographical area is essential for effective diagnosis and treatment of pollen allergy⁴. The relationships between allergic sensitization, allergen exposure and clinical allergic cross reactions play a vital role in elucidating the role of plant pollen allergens.

Profilin and polcalcin are known panallergens, accountable for a large number of allergic cross reactivity⁵. These are small acidic proteins with molecular mass of 14-15 kDa and 9 kDa, respectively. Major allergens from the Parthenium pollen have been referred to be in the range of 14-64 KDa⁶. Profilins are ubiquitous proteins found in all eukaryotic cells as an actin-binding protein that regulates microfilament dynamics and mediate changes in polymerization in the microtubule cytoskeleton; they are also implicated in multiple signaling pathways, including those involving the Ena/Vasodilator-stimulated phosphoprotein (VASP) family⁷. Profilins are highly conserved allergenic proteins present in many grasses, weeds, trees, fruits, vegetables and have also been identified in latex⁹,¹⁰.

On the other hand, polcalcins belong to a calcium binding protein family having two EF-hand sequential motifs, involved in the binding of Ca²⁺ ions and is

*Correspondence:
E-mail: jksingh@kluniversity.in
composed of 12 conserved amino acid residues. These are also implicated in signaling processes, neuronal exocytosis and pollen tube growth\textsuperscript{11}. Though, polcalcins are described as minor allergen, still they recruit a high percentage of specific IgE, ranging from 10 to 40% in allergic patients\textsuperscript{12,13}. Interestingly, IgE binding to these proteins depends on the presence of calcium levels in the allergic patients sera\textsuperscript{12,14}.

Both allergens are highly conserved among grass species and share a high degree of amino acid sequence identity with their counterparts from other allergenic sources (60-90% for polcalcins and >75% for profilins). This accounts for the high level of cross reactivity observed among members of the same family. The cross reactivity for profilin was first observed between the weed pollen profilins Art v 4 (mugwort) and Amb a 8 (ragweed)\textsuperscript{15}.

However, profilin has also been reported as a major allergen in pollen and food sources viz., Chenopodium album and Cucumis melo, respectively\textsuperscript{15,16} and in populations polysensitized to olive\textsuperscript{17,18} and grass\textsuperscript{19} pollen. Around 10-50% of allergic patients are sensitized to profilins\textsuperscript{9,20}. Irrespective of their designation as major or minor allergens, both profilin and polcalcin are important markers of polysensitization and at least in the case of profilins, previous sensitization to aeroallergens from pollens results in susceptibility to subsequent exposure to certain food counterparts, triggering oral allergy syndrome\textsuperscript{21}.

In India, Sorghum ranks third in area and production after Rice and Wheat. The crop accounts for nearly 52% of the area and 63% of production, under millets, with an area of 15.8 million hectares and a production of 11.85 million tons. Among the countries that grow Sorghum on commercial scale, India ranks 1\textsuperscript{st} in acreage and 2\textsuperscript{nd} in production. Maharashtra, Karnataka, Andhra Pradesh, Gujarat, Tamil Nadu and Madhya Pradesh are the major Sorghum growing states. Farmers exposed to Sorghum grain dust are more likely to experience respiratory symptoms such as cough and chest tightness as well as an acute febrile illness\textsuperscript{22,23}. A study conducted at Hyderabad, Andhra Pradesh has estimated allergenicity to Sorghum pollen grains to be 54.9\%\textsuperscript{24}. This analysis was reported with the crude extract of Sorghum pollens. However, the Sorghum pollen allergen proteins responsible for the allergy are not yet indentified. The present study is focused on identification and \textit{in silico} molecular characterization of SorbPF and SorbPC allergen epitopes.

Materials and Methods

\textit{Pollen collection—Sorghum bicolor} (L.) Moench (IS-10897) seeds were procured from ICRISAT (Hyderabad) and were grown in KL University fields. At the stage of flowering (earlier to mature), the inflorescence was collected into a sterile glass bottle and the pollen were segregated by tapping with the wooden stick onto the clean paper. After collecting all the pollen, the pure pollen were segregated by using the 80 and 160 mesh sieve.

\textit{RNA isolation and cDNA construction—}Total RNA was extracted by using Trizol reagent (Invitrogen). The RNA quality was gel checked and 1\µg of total RNA was used for cDNA construction using random primers following Invitrogen kit protocol.

\textit{Homology search—}Nucleotide sequences of Profilin (Cyd12 (Accession no.Y08390), Phlp12 (Accession no. X77583), Os12 (Accession no.AF310253), Arah5 (Accession no.JQ974980), Amba8 (Accession no.AY268426), Artv4 (Accession no.AJ421031), Betv2 (Accession no.M65179), Glym3 (Accession no.EU106621) and Polcalin (Phlp7 (Accession no.Y17835), Cynd7 (Accession no.U35683), Chea3 (Accession no.AY082338), Betv4 (Accession no.X87153), Alna4 (Accession no.Y17713), Olee3 (Accession no.AF015810), Syrv3 (Accession no.AF078681), Amba9 (Accession no.AY894657) and Artv5 (Accession no.AY904434) allergens were retrieved from NCBI (http://www.ncbi.nlm.nih.gov) databases. A homologous comparison was performed to determine their conserved regions by using ClustalW2 (EMBL-EBL) software. Two specific primers were designed based on the sequence of the conserved region. The sequences of the forward primer for Polcalcin-(F)-5'-ATGGCGGAGACGGCGGACATG-3' and reverse primer (R) 5'-TCAGAAGACCTTGGCGACGTC-3' and for Profilin—F—5'-ATGTCGTGGCA GACGTACGTC-3' and R-5'-TTACAGGCCCTT GCTCTACGAG-3' were used for PCR amplification.

\textit{PCR amplification—}The PCR was carried out in a total volume of 25 \µl, containing 1 \µl Taq-polymerase (1U), 2.5 \µl of 10X buffer, 10 mM dNTPs-1 \µl, 25 mM MgCl\textsubscript{2}-1 \µl, 10 pmol each forward and reverse primer, 1 \µg of cDNA and 16.5 \µl of Milli Q water. PCR variables were set as initial denaturation step at 94°C for 5 min followed by 35 cycles of 94°C for 30 s (denaturation), 60°C for 30 s (annealing) and 72°C for 1 min (renaturation) with a final extension at 72°C for
5min. The amplified PCR products were size-fractionated on 1.5% agarose gel.

Cloning, expression and purification of profilin and polcalcin—Full length coding sequences of *Sorghum* profilin and polcalcin were cloned into pET28b(+) vector (Novagen Inc., USA) and transformed into the *E. coli* BL-21 cells. The positive clones were confirmed by restriction digestion and sequencing. These were grown at 37°C to an OD of 0.4-0.6 at 600 nm and the expression was induced by adding 1 mM isopropyl-D-thiogalactopyranoside (IPTG) to the culture; the cells were harvested 4 h later. The harvested cells were frozen overnight at −20°C and resuspended in lysis buffer (20 mM Tris-HCl (pH 7.5), 300 mM NaCl, 5 mM imidazole and 1 mM PMSF) and sonicated to lyse the bacteria. The cell lysate was clarified by centrifuging at 15000 ×g for 15 min at 4°C; the supernatant was loaded onto a Ni-NTA column (Bio-rad) equilibrated with buffer containing 20 mM Tris-HCl (pH 7.5), 300 mM NaCl and 10 mM imidazole. After exhaustive washing with wash buffer (20 mM Tris-HCl (pH 7.5), 300 mM NaCl and 20 mM imidazole), the column was eluted with 200 mM imidazole in the equilibration buffer. The eluted protein was analyzed by SDS page stained with coomassie blue.

In silico analysis—The confirmed sequences were analyzed by using allergenicity testing software’s. EVALER (http://www.slv.se/en-gb/Group1/Food-Safety/e-Testing-of-protein-allergenicity/e-Test-allergenicity/) a web tool, where proteins can be tested electronically (e-Testing) to verify the potential affinity columns could be of potential use for diagnostic purpose.

In silico analysis—Multiple alignment was done for both *Sorghum* profilin and polcalcin proteins with known allergen polcalcin protein sequences of *Cynodon dactylon* (Cynd7), *Phleum pretense* (Phlp7), *Chenopodium album* (Chea3), *Betula verrucosa* (Betv4), *Alnus glutinosa* (Alng4), *Olea europaea* (Olee3), *Syringa vulgaris* (Syrv3), *Ambrosia artemisiifolia* (Amba9) and *Artemisia vulgaris* (Artv5); and profilin protein sequences of *Cynodon dactylon* (Cyd12), *Phleum pretense* (Phlp12), *Arachis hypogaea* (Arah5), *Glycine max* (Glyn3), *Ambrosia artemisiifolia* (Amba8), *Betula verrucosa* (Betv2) and *Artemisia vulgaris* (Artv4) (Fig. 2).

The SorbPC protein sequence showed a close similarity with the Cynd7 and Phlp7 allergen sequences where as SorbPF protein sequence showed close similarity to Phlp12, Os12 and Cynd12 allergen sequences (Fig. 2). Based on alignment studies, these proteins have shown more than 70% identity over the full-length peptide of the known allergens. It is, therefore, appealing to speculate that the *Sorghum* proteins reported in this study also hold allergenicity. In order to understand the evolutionary relationship between the SorbPC and SorbPF with other known allergens, phylogenetic analysis was done with all the homologous sequences using Tree top (http://www.genebee.msu.su/services/phtree_reduced.html) (Fig. 3).
Fig. 2—(A) SorbPC was aligned with known polcalcin allergen sequences; (B) SorbPF was aligned with the known profilin allergen sequences. [Symbol (*) indicates the exact match with other amino acids, (:) indicates same group amino acid change.]

Fig. 3—Molecular phylogenetic analysis by Maximum Likelihood method. (A) SorbPF protein sequence with known allergens. (The tree with the highest log likelihood (-1828.9282) is shown. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 9 amino acid sequences); (B) SorbPC protein sequence with known allergen sequences. (The tree with the highest log likelihood (-1002.0419) is shown). [The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 9 amino acid sequences. All positions containing gaps and missing data were eliminated. Evolutionary analysis was conducted in MEGA6.]
Further, identified SorbPC and SorbPF sequences were evaluated with Evalall tool (http://www.slv.se/en-gb/Group1/Food-Safety/e-Testing-of-protein-allergenicity/e-Test-allergenicity/) for e-Test allergenicity. Sorghum polcalcin showed 95.65, 92.31 and 68.75% homology to Cynd7, Phlp7 and Brassica rapa, respectively. The amino acids 30-52 position of Sorghum polcalcin was matching with both Cynd7 and Brassica rapa IgE binding site, whereas amino acids positioned at 55-80 were found homologous with Phlp7 IgE binding site.

Sorghum profilin did not show any homology to Cynd12 and phlp12, but interestingly showed 79.37, 79.59 and 76% homology to Litchi chinensis, Pho d 2 and Ambrosia artemisiifolia, respectively. The amino acids in the position 1-52 were homologous with Litchi chinensis, Pho d 2 and Ambrosia artemisiifolia profilin allergen proteins. For both the Sorghum proteins reported, Evalall assessment was presumably allergen and false alarm was 0%.

In addition, these proteins were also subjected to Algpred (http://www.imtech.res.in/raghava/algpred) screening to compare the allergenicity score of these allergens with other known allergens. The allergenicity score was 1.149 and 0.879 for SorbPF and SorbPC, respectively (Table 1).

<table>
<thead>
<tr>
<th>Gene</th>
<th>Length</th>
<th>Score</th>
<th>Threshold</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cynd12</td>
<td>131</td>
<td>1.392</td>
<td>-0.4</td>
<td>85.64</td>
<td>67.96</td>
</tr>
<tr>
<td>Phlp12</td>
<td>131</td>
<td>1.002</td>
<td>-0.4</td>
<td>85.64</td>
<td>67.96</td>
</tr>
<tr>
<td>SorbPF</td>
<td>111</td>
<td>1.149</td>
<td>-0.4</td>
<td>85.64</td>
<td>67.96</td>
</tr>
<tr>
<td>Os12</td>
<td>131</td>
<td>1.002</td>
<td>-0.4</td>
<td>85.64</td>
<td>67.96</td>
</tr>
<tr>
<td>Arah5</td>
<td>131</td>
<td>1.459</td>
<td>-0.4</td>
<td>85.64</td>
<td>67.96</td>
</tr>
<tr>
<td>Amba8</td>
<td>131</td>
<td>1.32</td>
<td>-0.4</td>
<td>85.64</td>
<td>67.96</td>
</tr>
<tr>
<td>Artv4</td>
<td>131</td>
<td>1.669</td>
<td>-0.4</td>
<td>85.64</td>
<td>67.96</td>
</tr>
<tr>
<td>Betv2</td>
<td>131</td>
<td>1.337</td>
<td>-0.4</td>
<td>85.64</td>
<td>67.96</td>
</tr>
<tr>
<td>Glym3</td>
<td>131</td>
<td>2.007</td>
<td>-0.4</td>
<td>85.64</td>
<td>67.96</td>
</tr>
</tbody>
</table>

(A) Sorb PF predicted allergenicity scores; and (B) Sorb PC predicted allergenicity score.

[The protein showing the score above 0.5 with a threshold-0.4 and PPV >80% is considered to be a potential allergen. PPV-Positive Prediction Value, NPV-Negative Prediction Value.]

Allergens from different sources are known to have an identical structural motif. Profilin has a ligand-binding site in the groove located inside a α-β motif and leads to the stimulation of T cell helper type 2 (Th2) responses and a subsequent bias towards the synthesis of IgE. Protein modeling studies showed >80% structural relation to Betv2 profilin allergen. Comparison of Phlp12 IgE epitope (LGHDGTWVAQSADF) with SorbPF showed only a single amino acid change from L to I at 26th amino acid position. This clearly indicates allergic patients to Phlp12 profilin allergen would be cross reactive with SorbPF allergen.

Calcium binding proteins have two or more EF-hand sequential motifs, which are implicated in the binding of Ca2+ ions and are composed of 12 conserved amino acid residues. Binding of the allergenic proteins to IgE antibodies of allergic patients' sera is dependent on the presence of calcium. A motif search (DTNGDGKISLSEI and DTGDGFIDFNEF) revealed the presence of two typical calcium binding domains (EF-hands) in the deduced SorbPC amino acid sequence exactly matching to the Phlp7 allergen sequence. Comparison of amino acid sequence with other pollen polcalcin allergens confirmed best homology with a group of low molecular weight, two EF-hand calcium binding pollen allergens of Aln g 4 and Cynd 7. SorbPC exhibits high homology with Phlp7 and Cynd7 and would also exhibit cross reactivity to the allergen patients IgE.

This work is the first report of the SorbPC and SorbPF coding sequences and in silico identification of the allergenic epitopes on these proteins. Subsequently, recombinant expression and His-tag purification of allergens is reported. Further, the role of allergenic epitopes in eliciting the allergic response in Indian population needs to be studied and can lead to selection of right epitope peptides for the development of diagnostic kits for allergenicity to Sorghum pollen grains.

Acknowledgements
The authors acknowledge the funding from University Grants Commission (UGC), New Delhi, (UGC-MRP F. 34-260/2008(SR). Author JSB thank UGC for the Research Award (No. F. 30-1/2013 (SA-II)/RA-2012-14-GE-ANP-1237).

References


