Protein Profiles of Indigenous and Commercial Extracts of *Amaranthus* Pollen for the Diagnosis of Allergy and Asthma Patients

Ayodele A. Alaiya, Halima A. Alsini, Mohammed O. Gad El-Rab and Syed M. Hasnain

King Faisal Specialist Hospital and Research Centre
College of Medicine, King Saud University, Riyadh, Saudi Arabia

**Abstract:** Background: Pollen grains from *Amaranthus viridis* (Av) (Slender amaranth, pigweed, etc.) are known to be allergenic and a potential cause of upper and lower respiratory allergic diseases, but neither A.v. extract nor its pollen grains are available commercially. Materials and Method. *Amaranthus* spp extracts were prepared from pollen grains of both indigenous and commercial species and SPT was conducted on allergic patients. Protein patterns of seven different types of *Amaranthus* samples as well as serum samples from patients were analyzed by label-free expression proteomics using liquid chromatography coupled with tandem mass spectrometry (LC/MS/MS). Results: Seven (7) proteins that were significantly differentially expressed and distinctively discriminate between patient and control samples, were identified. Only immunoglobulin heavy constant gamma 2 showed greater than 2-fold higher mean expressions in the patient samples than control the observed differences were significant (p < 0.0001). Two of the proteins Ig gamma-2 chain C region and Carboxypeptidase N, polypeptide 2 are potential surrogate markers of allergy and asthma. The antigenic variations within *Amaranthus* species are likely to identify a precise diagnostic species for selection and choice of extracts for some parts of the world. Conclusion: This study highlights the advantage of proteome analysis of different allergens towards discovery of surrogate biomarkers that may become potentially useful to complement currently existing tests for Allergy and Asthma patients.

**Abbreviations:** SPT (Skin Prick Test) • Av (*Amaranthus viridis*) LC/MS/MS (liquid chromatography coupled with tandem mass spectrometry).

**Key words:** Allergens • Amaranthus • Protein • Stimulating biomarkers

**INTRODUCTION**

Bronchial Asthma is a common allergic disease occurring in all age groups, particularly in children and the prevalence in the world is increasing [1]. In a study conducted in KSA, an overall increase in the prevalence of asthma was recorded from 8% to 23% over a period of 9 years, representing a 3 fold increase [2]. Wheezing constituted 11.5% with a consequent increase in morbidity and mortality [3-5].

The prevalence of allergic rhinitis (AR) is also increasing worldwide. Globally, over 400 million people suffer from AR [6]. The most common cause of AR is sensitivity to airborne allergens. The nature and number of aeroallergens differ according to geographical variability, vegetation parameters, climatic and seasonal changes and locality even within the same broader region [7].

Exposure to aeroallergens has long been associated with airway allergic disorders. In recent decades a number of authors have argued that allergen exposure is the major primary cause of asthma [8-11] and that the global increases in asthma prevalence could be the result of increases in exposure to aeroallergens [10, 12].

There are a few *Amaranthus* species in Saudi Arabia, but the dominant species on the ground and frequently encountered pollen in the air belongs to *A. viridis* and *A. lividus* [13].

Contrary to the cross reactivity, treatment by immunotherapy may not be successful unless precise molecular relation between offending allergen and desensitizing allergens are established.

While clinical diagnosis of most allergy conditions can be accurately made, specific geographical implicated allergens and predictions of immunotherapy treatment response elude the currently available diagnostic and
therapeutic tools for care of allergy patients across different regions of the world. Therefore, there is a need for the discovery of sensitive and specific biomarkers for accurate disease diagnosis, prognosis and predicting treatment response. This study aimed to identify specific allergy stimulating (in patients or control) proteins that might be produced in response to different allergens (indigenous and commercial spp) and released as circulating products, which are detectable in serum using Expression Proteomics approach.

MATERIALS AND METHODS

*Amaranthus* pollen (*A viridis* and *A. lividus*) were collected indigenously. Commercial pollen were purchased from: (Greer Laboratory, USA): *Amaranthus palmeri*, *Amaranthus tuberculatus*, *Amaranthus retroflexus*, *Amaranthus hybridus* and (Allergon Company, Europe): *Amaranthus retroflexus*, *Amaranthus tamariscinus*.

Extracts were prepared as mentioned [14]. Protein content of each extract was determined by Bradford method [15].

Skin prick test (SPT) was performed on 17 allergic patients, 10 healthy non allergic subjects, Phosphate buffered saline and histamine was negative and positive control respectively.

Venous blood was drawn from patients and sera was separated and stored at -20°C. Blood samples from 10 healthy volunteers were also collected to act as control.

Protein extracts for 2D SDS-PAGE: Phenol extraction followed by methanolic ammonium acetate precipitation [16].

**Dimensional (2D) Electrophoresis:** Pellet samples from eight different *Amaranthus* species were dissolved in lysis buffer and First-dimension isoelectric focusing was carried out for a total of 45, 500 Vh in a PROTEAN IEF cell (Bio-Rad). IPG-strips were then loaded and run on a 12.5% SDS-PAGE and run for 2 hours at 200V. Crude serum samples were diluted to a total volume of 350 µl and the same protocol was applied for the 2D running as mentioned before [17, 18].

Gels were stained with silver nitrate; scanned using a calibrated densitometer, GS 800 and data was analyzed using Progenesis SameSpots software (Nonlinear Dynamics) and PDQUEST (Bio-Rad) as previously described [19, 20].

Proteomic Analysis, Sample Preparation, Protein in Solution-Digestion and LC-MS Analysis: For each analysis sample group, 200µg complex protein mixture was taken and exchanged twice with 500 µL of 0.1% RapiGest (Waters, Manchester, UK) (1 vial diluted in 1000 µL 50 mM Ammonium-bicarbonate) with a 3-kDa ultra filtration device (Millipore). Protein concentration of between 0.1 and 1 µg/µL was achieved at the end of digestion. Proteins were denatured in 0.1% Rapi Gest SF at 80°C for 15 minutes, reduced in 10 mM DTT at 60°C for 30 min, spin down condensate and allowed to cool to room temperature and alkylated in 10 mModoacetamide for 40 min at room temperature in the dark. Samples were trypsin digested at a 1:50 (w/w, (1µg/µl trypsin concentration)) enzyme: protein ratio and trypsin, overnight at 37°C. The reaction was quenched with 4µl of 12M HCl at 37°C for 15 min and centrifuged at 13000 RPM for 10 min. All samples were spiked with yeast alcohol dehydrogenase (ADH; P00330) as internal standard to the digests to give 200 fmol per injection for absolute quantitation prior to LC/MS analysis.

Protein Identification by Mass Spectrometry -LC/MSE:
Expression proteomics for biomarker discovery of both qualitative and quantitative protein changes was performed using 1-Dimensional NanoAcquity liquid chromatography coupled with tandem mass spectrometry on Synapt G2. All analyses were done on Trizaic Nano source (Waters, Manchester, UK) ionization in the positive ion mode nanoESI.

A total of 2 µg protein digests was loaded on-column and samples were infused using the Acquity sample manager with mobile phase consisting of A1 95% (water + 0.1% Formic acid) and B1 5% (Acetonitrile + 0.1% formic acid) with sample flow rate of 0.500 µl/min. Data-independent acquisition (MSE) / iron mobility separation experiments were performed and data was acquired over a range of m/z 50 – 2000 Da with a scan time of 1 sec, ramped transfer collision energy 20 -50 V with a total acquisition time of 120 min. All samples were analyzed in triplicate runs and data were acquired using the Mass Lynx programs (version. 4.1, SCN833, Waters, Manchester, UK) operated in resolution and positive polarity modes. The acquired MS data were background subtracted, smoothed and de-isotoped at medium threshold. Protein Lynx Global Server (PLGS) 2.2 (Waters, Manchester, UK) was used for all automated data processing and database searching. The generated
peptide masses were searched against Uniprot protein sequence database using the PLGS 2.2 for protein identification (Waters, UK).

**Data Analysis and Informatics:** TransOmics Informatics (Waters Corporation, UK) was used to process and search the data. A Human database containing 46906 reviewed entries was downloaded from Uniprot. A decoy database was created by reversing this and concatenated to the original database prior to searching. The principle of the search algorithm is described [21]. The following criteria were used for the search: 1 missed cleavage, Max protein mass 1000kDa, Trypsin, Carbamidomethyl C fixed and Oxidation M variable modifications.

The data was filtered to show only statistically (ANOVA) significantly regulated proteins ($p \leq 0.05$) with $\geq 3$ peptides ID and a fold change $>2$.

Additionally, ‘Hi3’ absolute quantification was performed using ADH as an internal standard to give an absolute amount of each identified protein (Waters Corporation, UK).

**RESULTS AND DISCUSSION**

**Protein Expression Patterns to Different Species of *Amaranthus* Allergens:** The SPT reaction using different *Amaranthus* allergens as measured by the wheal diameter (mm) was evaluated in all 17 patients. Interestingly, we observed unique patterns in the wheal diameters between the indigenous and commercial extracts. The two local indigenous *A. viridis* and *A. lividus* demonstrated very similar expression patterns compared to the 5 other commercial extracts. Majority of patients (64%) showed low reactivity to both *A. viridis* and *A. lividus* and only 36% of patients are highly reactive to the two indigenous *Amaranthus* allergens. On the other control, all the 5 commercial allergens were highly expressed in majority of the patient (approx. 61.5%) and only 38.5% of patients showed low reactions to the entire commercial extracts. Even though there is a clear clustering of patient and control samples based on their protein expression patterns, we did not see any clear clustering among the patients that reacted differently to the different *Amaranthus* species.

**Protein Expression Profiles Between Control and Sub Groups of Patients Responding to Indigenous *A. Lividus* and *A. Viridis:** Ninety one 91 protein spots were significantly differentially expressed between control and sub groups of patients exposed to indigenous

* A. lividus and *A. viridis* only 22 of the 91 protein spots were highly expressed in patients showing low expression to *A. viridis*+ *A. lividus*, while 20 protein spots were highly expressed in patients with high reaction to *A. viridis*+ *A. lividus*. The 91 dataset resulted in clear clustering of all control samples and two clusters of patient samples demonstrating low and high expressions to these protein spots (Figures 1 & 2).

**Protein Expressed Profiles Between Control and Sub Groups of Patients Responding to Commercial Am Allergens:** When similar analysis described above was done for patient’s response to all commercial allergens, a total of 84 protein spots were significantly differentially expressed between control and sub groups of patients exposed to all the 5 commercially purchased *Amaranthus* allergens. The 84 dataset resulted in clear clustering of all control samples and two clusters of patient samples demonstrating low and high expressions to these protein spots.

**Heterogeneity in Patient Response to Indigenous and Commercial Allergens:** The 91 protein spots that were significantly differentially expressed between control and sub groups of patients exposed to indigenous *A. lividus* and *A. viridis* and the corresponding 84 differentially expressed protein spots among commercial allergens were compared. Only 18 protein spots falls in the intersection of the two datasets. This indicates high degree of heterogeneity in the patient response to indigenous and commercial allergens. This then call for efforts in incorporating indigenous allergens in future production of commercial allergens extracts.

**Global Protein Expression Profiles of Serum Samples:** Serum samples obtained from 13 patients (with closely normalized protein concentrations) of the 17 patients diagnosed using different *Amaranthus* allergens as well as sera from 10 healthy subjects as negative control samples were analyzed. Crude serum samples were prepared and analyzed by label-free liquid chromatography coupled with tandem mass spectrometry for both qualitative and quantitative differences in the expression of multiple polypeptides.

**Differentially Expressed Protein Features:** The 91 identified proteins from patient samples and control samples were analyzed. Seven of these proteins were differentially expressed significantly with more than 2 fold change in their expression levels between sample groups.
Fig. 1: The Principal Component Analysis Plot of Expression profiles between control and sub groups of patients responding to commercial Amaranthus allergens samples using the expression dataset of 91 polypeptides. (The samples are classified as pink, Control, Purple, - patients with low commercial Am reactions and Pink, - patients with high commercial Am reactions). Statistically differentially expressed proteins were used for the Principal Component Analysis (ANOVA, p < 0.0001 and = 2 fold-changes in expression levels). The plot was generated using ProgenesisSameSpots 2-DE analysis software program version 4.1 (Nonlinear Dynamics).

Fig. 2: The Principal Component Analysis Plot with a distinct classification of control and sub groups of patients responding to indigenous *A. lividus* and *A. viridis* using the expression dataset of 91 polypeptides. (The samples are classified as pink, Control, Purple, - patients with high Am V + Am Lv reactions and Pink, - patients with low Am V + Am Lv reactions). Statistically differentially expressed proteins were used for the Principal Component Analysis (ANOVA, p < 0.0001 and = 2 fold-changes in expression levels). The plot was generated using ProgenesisSameSpots 2-DE analysis software program version 4.1 (Nonlinear Dynamics).

A data set of the 7 proteins clearly discriminates the samples into 2 distinct groups by unsupervised Hierarchical Cluster analysis and correspondence analysis (Figure 3A& B).

**Functional Interpretation of the Identified Proteins:**
Further characterizations of the identified proteins were explored using Ingenuity pathway analysis (Ingenuity Systems, Inc.CA, USA). The 7 identified proteins were
Fig. 3: (A) Hierarchical Cluster analysis using the seven (7) identified proteins with significant change between patients and control subjects. The names of the identified proteins are indicated in the dendrogram (Red, patients and Blue, control subjects). The dendrogram was generated using the Bray Curtis distance metric and an average linkage clustering method from the J-Express software. (B) Same dataset was subjected to Correspondence Analysis (CA) and the expression changes allow clear separation into two distinct groups.

Table 1: Summarized functional characteristics of some of the identified proteins.

<table>
<thead>
<tr>
<th>Accession</th>
<th>Peptides used for quantitation</th>
<th>Confidence score</th>
<th>Anova (p)</th>
<th>Max fold change</th>
<th>Highest mean condition</th>
<th>Lowest mean condition</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>P2742,P2742-2</td>
<td>11</td>
<td>126.8224</td>
<td>5.82E-06</td>
<td>5.41181778</td>
<td>CTRL</td>
<td>Patient</td>
<td>Pregnancy zone protein OS=Homo sapiens</td>
</tr>
<tr>
<td>P01859</td>
<td>2</td>
<td>55.2519</td>
<td>0.0001529</td>
<td>2.30426643</td>
<td>CTRL</td>
<td>Patient</td>
<td>Ig gamma-2 chain C region OS=Homo sapiens</td>
</tr>
<tr>
<td>P22792</td>
<td>1</td>
<td>6.3086</td>
<td>0.00050263</td>
<td>2.07447515</td>
<td>CTRL</td>
<td>Patient</td>
<td>Carboxypeptidase N subunit 2 OS=Homo sapiens</td>
</tr>
<tr>
<td>P04278,P04278-2</td>
<td>2</td>
<td>15.4397</td>
<td>0.00055677</td>
<td>8.36336578</td>
<td>CTRL</td>
<td>Patient</td>
<td>Sex hormone-binding globulin OS=Homo sapiens</td>
</tr>
<tr>
<td>P01009,P01009-2</td>
<td>17</td>
<td>169.5696</td>
<td>0.00567003</td>
<td>2.22846038</td>
<td>CTRL</td>
<td>Patient</td>
<td>Alpha-1-antitrypsin OS=Homo sapiens</td>
</tr>
<tr>
<td>P02774,P02774-2</td>
<td>7</td>
<td>69.2711</td>
<td>0.00716535</td>
<td>3.78329766</td>
<td>CTRL</td>
<td>Patient</td>
<td>Vitamin D-binding protein OS=Homo sapiens</td>
</tr>
<tr>
<td>P00751,P00751-2</td>
<td>4</td>
<td>37.5715</td>
<td>0.03351649</td>
<td>2.3159042</td>
<td>CTRL</td>
<td>Patient</td>
<td>Complement factor B OS=Homo sapiens</td>
</tr>
</tbody>
</table>

Ctrl, Control, p, p value as expressed as level of significance

mapped and only two of them (carboxypeptidase N, polypeptide 2 and immunoglobulin heavy constant gamma 2) were represented in two (2) sub-signaling networks.

The two proteins were implicated and representation in a number of canonical pathways that includes allograft rejection signaling; autoimmune disease signaling; communication between innate and adaptive immune cells; primary immunodeficiency signaling; role of macrophages, fibroblasts and endothelial cells in rheumatoid arthritis and systemic lupus erythematosus signaling. The network in which the molecules are represented is as illustrated in Figure 4.

The functional domains of the majority of the identified proteins acts as peptidases and resides in subcellular location, extracellular space and plasma and they are commonly expressed and localized in blood, plasma/serum and urine. (Table 1 was partly generated from the ingenuity pathway analysis program).

All but one of the 7 identified differentially expressed proteins was highly expressed among the control samples. Only immunoglobulin heavy constant gamma 2 showed greater than 2-fold higher mean expressions in the patient samples than control, the observed differences were significant p< 0.0001) (Table 1).

All but one of the 7 identified differentially expressed proteins was highly expressed among the control samples. Only immunoglobulin heavy constant gamma 2 showed greater than 2-fold higher mean expressions in the patient samples than control, the observed differences were significant p< 0.0001) (Table 1).

Studies have suggested that a wide range of allergens and anti-allergy agents exert their actions through induction of mast cells and cytokines mediated reactions. Sequence of events occurs as the body reacts to different foreign particles such as allergens [22].

This study was done in an effort to identify specific allergy stimulating molecules that might be produced and released as circulating products, which are detectable in serum.

Biomarkers are often defined as specific enzymes or proteins that can be quantitatively measured and evaluated as objective indicators of normal biological, pathological, as well as therapeutic responses.
processes. Successful completion of the human genome project has accelerated advancements in proteomic technologies that lead to huge interest in translational research. Proteomics studies have resulted in identification of disease related or tissue specific proteins that could be potentially useful as disease biomarkers [23, 24].

We have identified 7 proteins and their expression pattern distinctively discriminate between patients and control subjects as shown in Figure 3A and B.

The majority of the proteins in the control group showed a marked altered abundance of greater than 2-folds up regulated than in the patient group. Only one protein (Ig gamma-2 chain C region) was found to be up regulated in the patient group, indicating that presence of these proteins might provide protection against allergens or as potential as therapeutic targets.

The functional characteristics of the identified proteins were further explored using the Ingenuity pathway analysis. Only 2 of the 7 identified proteins were implicated in different signaling networks including allograft rejection signaling; autoimmune disease signaling; communication between innate and adaptive immune cells; primary immunodeficiency signaling; role of macrophages, fibroblasts and endothelial cells in rheumatoid arthritis and systemic lupus erythematosus signaling. The summarized functional characteristics of some of the identified proteins were listed in Table 1. The data was partly obtained using the Ingenuity Pathway Analysis program.

Carboxypeptidase N, polypeptide 2 (ACBP/CPN2) is a glycoprotein that is synthesized by the liver and secreted into the plasma. It comprises of 2 identical 83-kD regulatory subunits as well as 2 identical 50-kD catalytic subunits. The protein is often otherwise called serum or plasma carboxypeptidase B protein.

These potent pro-inflammatory peptides have been suggested to play a key role in the pathogenesis of the allergic responses [22].

Carboxypeptidases (CPs), such as carboxypeptidase N (CPN) (kininase I, might be involved in the regulation of peptide-mediated vasodilation and respiratory mucosa vascular permeability via pro-inflammatory peptides such as bradykinin, anaphylatoxins and neuropeptides.
especially during allergic and non-allergic inflammation [25]. Their results were suggestive that plasma is the predominant source of secreted CP activity in human nasal mucosa, while plasma extravasation and interstitial fluid exudation across the epithelial accounts primarily for its presence in nasal secretions.

We observed greater than 2-fold high expression of Carboxypeptidase N, polypeptide 2 among the control samples than in patients sera. This molecule is involved in the metabolism of bradykinin peptide.

The peptide is known to be broken down in humans, by three kininases including angiotensin-converting enzyme (ACE), amino peptidase P (APP) and carboxypeptidase N (CPN) [26].

Studies have demonstrated bradykinin and other inflammatory response peptides are generated in nasal secretions upon nasal challenge of allergic individuals with appropriate allergen and have suggested that these potent pro-inflammatory peptides may contribute to the pathogenesis of the allergic response [24, 27]. Our observed decrease expression of Carboxypeptidase N, polypeptide 2 in this study agrees with other findings that in response to allergens, the protein is depleted due to its involvement in the metabolism of bradykinin. Our observation of high expression among the control non-allergic, individuals than in the allergic patients calls for further study as potential target as therapeutic agent or marker for monitoring treatment response in allergic patients.

In a biomarker discovery study using the classical 2-dimensional gel electrophoresis revealed that while psoriasin was significantly down regulated, Ig gamma-2 was highly expressed among allergic rhinitis patient along with other proteins that were significantly differentially expressed [28]. We observed >2-folds greater expression level of Ig gamma-2 among all the allergic patients than control non-allergic individuals. This observation might be potentially useful as complementary diagnostic or novel therapeutic target or as marker to monitor treatment response in allergic patients.

In summary, we have used expression proteomics for discovery of potential surrogate protein biomarkers that demonstrated diagnostic/discriminatory ability between all allergic and non-allergic samples.

**ACKNOWLEDGEMENTS**

This project was supported by King Abdul Aziz City for Science and Technology (KACST) Saudi Arabia under grant ARP# 27-11 and approved by King Faisal Specialist Hospital and Research Centre (RAC 2050 029).

The authors also wish to acknowledge Dr. Jonathan Fox at Waters, Manchester UK for technical assistance with Synapt G2 system and Mr. Mustafa Adam at King Khalid University Hospital for technical assistance.

**REFERENCES**


