Available online on www.ijppr.com
International Journal of Pharmacognosy and Phytochemical Research 2016; 8(5); 827-830

ISSN: 0975-4873

Research Article

Purification and Identification of 20kDa Protein from *Parthenium hysterophorus*

Praveen Kumar Vemuri[1]*, Bhavana Talluri[1], Gayathri Panangipalli[1], Siva Kumar Kadiyala[1], Suryanarayana Veeravalli[2], Vijaya Lakshmi Bodiga[3]

[1]*Centre for Genomics and Proteomics, Department of Biotechnology, K L University, Vaddeswaram, Guntur District, Andhra Pradesh, India*
[2]*Department of Humanities and Basic Sciences, Aditya Engineering College, Kakinada, East Godavari District, Andhra Pradesh, India*
[3]*Department of Molecular Biology, Institute of Genetics & Hospital for Genetic Diseases, Osmania University, Hyderabad, Telangana, India*

Available Online: 1st May, 2016

ABSTRACT
The study was aimed to investigate which orthologues of *Parthenium hysterophorus* genes might encode the P20 protein. Polyethylene glycol fractionation of *Parthenium hysterophorus* leaf proteins has shown detection of low-abundance proteins. Samples incubated at -80°C also reduced precipitation time of protein to two hours. Further, sonication was used for removal of non-protein contaminants which increased protein solubility and improved protein identification by MALDI-TOF mass spectroscopy. Tryptic digest of purified protein revealed the predicted size of the protein is ~20kDa by MALDI-TOF and identification of peptides present in the P20 protein is done by peptide sequence analysis. This study describes how a combination of bioinformatics and proteomics approaches led to the identification of novel P20 candidates.

Keywords: *Parthenium hysterophorus*, P20, Bioinformatics, MALDI-TOF, Proteomics.

INTRODUCTION

Bioinformatics is a useful tool in studying gene characterisation and function using information technology and covers a wide range of applications including sequence analysis, genome annotation, gene and protein expression analysis, protein structure prediction and many modeling techniques. Databases allow access to readily available data and also the opportunity to submit new data as it becomes available. GenBank[1] is a DNA sequence database that contains sequences submitted from individual laboratories and from data exchange from other international sequence databases, from many different species[2]. This database will be used to select *P. hysterophorus* leaf protein candidates based on their expression profiles. Proteomics is the large-scale analysis of proteins in living cells which can be used to identify proteins and to characterise protein expression, localisation, activity, regulation and post-transcriptional modification. One of the major techniques utilised for proteomics analysis is mass spectrometry (MS). Mass spectrometry techniques have been used in a number of plant species (including *A. thaliana*, pea and tobacco) to investigate the proteomes of mitochondria[3], chloroplasts[4], cell walls[5], vacuoles[6], nuclei[7] and specifically in pollen[8]. Here, bioinformatics and proteomic studies[9] - were used to identify the *P. hysterophorus* leaf protein that corresponds to P20.

MATERIAL AND METHODS

Extraction of plant pollen proteins

Whole plants were collected from K L University campus, Andhra Pradesh, India and made into smoothie[10] with 12% of polyethylene glycol at low, medium and high speed for 30 seconds each. The smoothie was passed in Microcon centrifugal filters (MRCFOR30) for high recovery of protein with concentration factor of <10X. Protein concentration was determined following the BCA method using a Protein Assay Kit (Thermo Pierce -23255).

SDS-PAGE analysis of purified protein

A volume of 10ul was added to loading buffer (Merck Biosciences) and incubated at 95^0C for 5 minutes. Samples were loaded along with medium range ready protein marker (Puregene) and electrophoresis was run for 2 hours at 100V or until the gel loading dye reached to the end of gel. Gel was washed and fixed in 50% methanol solution for few minutes and stained with Ezee blue direct stainer (Merck Biosciences) for 40 minutes. After staining, gel was imaged using gel doc (UVI-Tech) and or analyzed by white illuminator[11].

Proteomics approach to P20

In conjunction with the bioinformatics analysis, attempts were made to gain additional information as to the identity of the P20 protein directly, using a proteomics approach. A total protein extract of timothy grass (as control) along with *P.hysterophorus* was prepared and fractionated on a

*Author for Correspondence

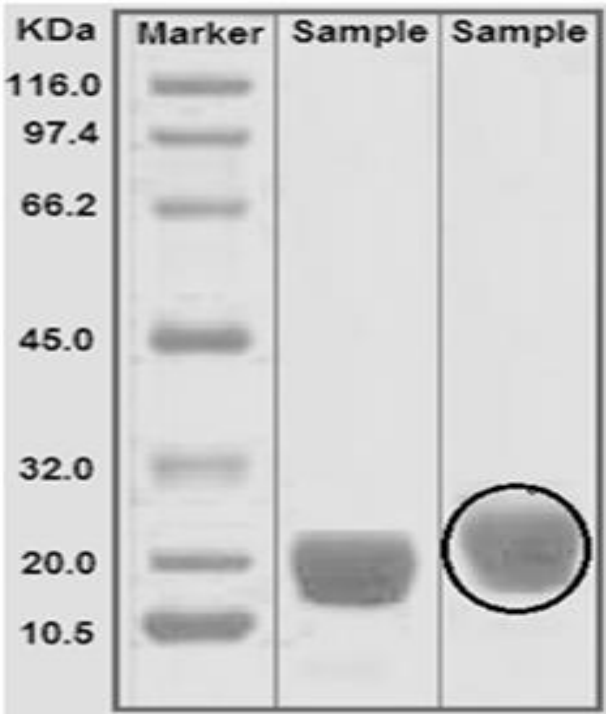

Figure 1: Ezee blue-stained gel of P20 protein on 12% SDS-PAGE extracted from *P. hysteroporus* leaves.

Table 1: Expected and observed average molecular masses of tryptic peptide fragments of P20 are shown with their corresponding position within the protein, expected and observed molecular mass, number of missed cleavages, and amino acid sequences.

Fragment	Expected Average Molecular Mass	Observed Average Molecular Mass	Missed cleavages
1	1526.56	1322.99	0
2	1652.32	1410.41	0
3	1598.02	1499.27	0
4	1593.56	1591.58	0

12% SDS-PAGE gel. Ezee blue stain was used to detect the separated proteins and a protein band corresponding to p20 was excised. The protein band was digested with trypsin, a serine endopeptidase that catalyzes the hydrolysis of peptide bonds on the carboxyl side of arginine and lysine residues, to digest the proteins into smaller fragments for analysis. These peptide fragments were then sent for analysis via MALDI-TOF MS (MS/MS) to determine the amino acids present (data not shown), from the mass of the peptides[12].

MALDI-TOF and Amino terminal analysis of P20

Protein band corresponding to 20 kDa was excised from the gels, digested with trypsin[13] and processed for mass spectrometric fingerprinting. In brief, peptide mixtures were partially fractionated on Poros 50 R2 RP microtips and the resulting peptide pools were analyzed by MALDI Biotyper (Brüker Franzen, USA) to enhance performance, simplify operation. Selected mass values were then taken to search a protein non-redundant database (NR; National Center for Biotechnology Information) using the Mascot Peptide Search algorithm[14].

Sequence alignment and phylogenetic tree construction of P20

Sequence alignment was performed using ClustalW2 to calculate the best match for the selected sequences, and lines them up to generate phylogenetic tree for retrieved protein families[15]. Significance for the modes was estimated using the protein weight matrix (gonnet as default value) and the alignments were adjusted using Bioedit V7.2

RESULTS AND DISCUSSION

SDS-PAGE analysis of purified protein

To determine the levels of P20 protein, samples from concentrated tubes were loaded in 12% polyacrylamide gel. Gels were incubated in fixing solution (50% methanol) for 30 min with two exchanges, washed three times with deionised water 10 min each and stained in Ezee blue direct stainer solution for overnight or were stored in the staining solution until the bands of interests were visualized. In lane M, Molecular marker procured from Puregene was used for the determination of protein molecular mass. While in Lane 2 and 3 the purified P20 from *P.hysteroporus* was loaded to check the purity. The molecular weight of protein was about 20KDa; we obtained relatively pure P20 that gave a considerable yield. The gels were scanned on a Gel scanner with white light converter (UVI-Tech, Lark Innovative) and the resulting images were analyzed with UVI-Tech Software (Figure 1)

MALDI-TOF analysis and N-terminal sequencing of P20

The nature of the differences between expected and observed masses of the purified P20 was investigated by trypsin digestion and MALDI-TOF mass spectrometry for the purpose of mass spectrometric fingerprinting as done earlier[16]. Ionization spectrum for the masses of peptides liberated by trypsin digestion shows four most prominent peaks; the corresponding *m/z* values shown in table 1, were taken to query the National Center for Biotechnology Information (NCBI) non-redundant protein sequence database for pattern matches, using the Mascot Peptide Search program. The resulting masses were compared with the expected peptide masses and amino acid sequences obtained after *in silico* digestion.

Sequence alignment and phylogenetic tree construction of P20

EMBOSS Backtranseq back-translates protein sequences to nucleotide sequences was used to predict the gene sequence of the allergic proteins. The amino acid sequence of the peptide was entered as input sequence and the codon table usage table was selected as *Arabidopsis thaliana* as control. The results were reported for all the possible peptides generated from ionization spectrum of P20 tryptic digest. As shown in Figure 2, we first retrieved all the pollen gene sequences using as the major molecular consensus defining the entire super family of pollen proteins. The number of pollen genes greatly varied from one plant species to another. At present, more than half of the catalogued plant pollen protein families encoded a single pollen-like gene, which was in most cases

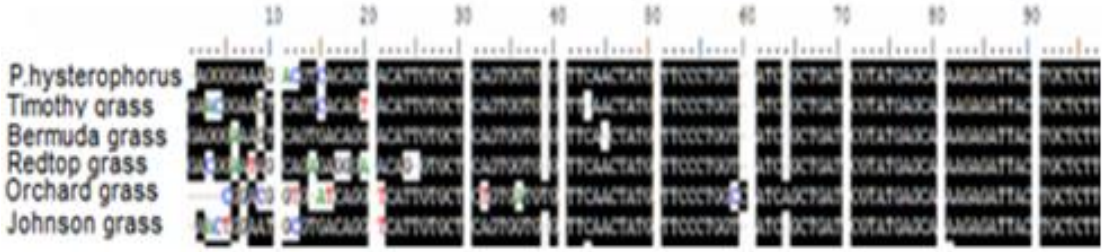

Figure 2: Alignments of novel genomic DNA sequences of partial sequences of plant allergenic gene from P. hysterophorus, Timothy grass, Bermuda grass, Redtop grass, Orchard grass and Johnson grass by using ClustalW Multiple Alignment. The alignment shows that this specific fragment of allergic gene is highly conserved among these plants.

Figure 3: Phylogenetic tree of partial allergenic gene sequences isolated from *P. hysterophorus*, Timothy grass, Bermuda grass, Redtop grass, Orchard grass and Johnson grass by using UPGMA software. The tree shows that allergenic genes of these plants have close evolutionary relatedness.

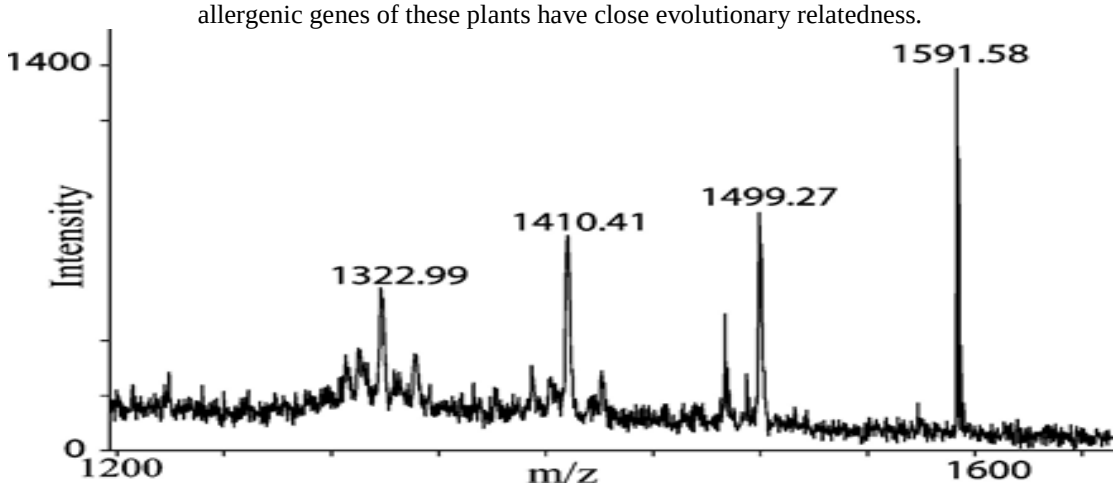

Figure 4: MALDI-TOF peptide mass fingerprint spectrum of trypsinized P20 protein.

"uncharacterized". As shown in Figure 3, phylogenetic tree of partial allergenic gene sequences isolated from *P. hysterophorus*, Timothy grass, Bermuda grass, Redtop grass, Orchard grass and Johnson grass by using UPGMA software is constructed. The tree shows that allergenic genes of these plants have close evolutionary relatedness.

CONCLUSION

In this study, the purpose was to characterise the partial portion of *Parthenium hysterophorus* leaf protein and its sequence by using bioinformatics tools and compare its homology with other known allergic proteins. We isolated a 20-kDa protein from *Parthenium hysterophorus* leaves that shows allergic reactions. This protein shares little amino acid sequence homology with any other proteins, including proteins from Timothy grass[17], Bermuda grass[18], Redtop grass[19], Orchard grass[20] and Johnson grass[21]. These novel partial fragments of pollen genes from these wild medicinal plants can be used as internal controls for future gene expression studies of these important plants after precise validations of their stable expression in such plants. This is the first report on identification and characterization of such internal control gene for expression studies among variety of wild plants that possesses economical and medicinal values. Thus, it

constitutes a new class of protein but may require many other methods to be investigated likely for the expression of allergic characteristics of the plant.

ACKNOWLEDGEMENT

The authors gratefully thank the management of K L University for providing the support and encouragement for conducting this work and special thanks are extended to the laboratory assistants during this work.

REFERENCES

1. Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Wheeler, D.L. (2008) GenBank. *Nucleic acids research* 36, D25-D30.
2. Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Sayers, E.W. (2012) GenBank. *Nucleic acids research* gks1195.
3. Brugière, S., Kowalski, S., Ferro, M., Seigneurin-Berny, D., Miras, S., Salvi, D., and Joyard, J. (2004) The hydrophobic proteome of mitochondrial membranes from Arabidopsis cell suspensions. *Phytochemistry* 65, 1693-1707.
4. Gómez, S.M., Nishio, J.N., Faull, K. F., and Whitelegge, J.P. (2002) The chloroplast grana proteome defined by intact mass measurements from liquid chromatography mass spectrometry. *Molecular and Cellular Proteomics* 1, 46-59.
5. Feiz, L., Irshad, M., Pont-Lezica, R.F., Canut, H., and Jamet, E. (2006) Evaluation of cell wall preparations for proteomics: a new procedure for purifying cell walls from Arabidopsis hypocotyls. *Plant Methods* 2, 1.
6. Carter, C., Pan, S., Zouhar, J., Avila, E.L., Girke, T., and Raikhel, N.V. (2004) The vegetative vacuole proteome of Arabidopsis thaliana reveals predicted and unexpected proteins. *The Plant Cell* 16, 3285-3303.
7. Pendle, A.F., Clark, G.P., Boon, R., Lewandowska, D., Lam, Y.W., Andersen, J., and Shaw, P.J. (2005) Proteomic analysis of the Arabidopsis nucleolus suggests novel nucleolar functions. *Molecular biology of the cell* 16, 260-269.
8. Holmes-Davis, R., Tanaka, C.K., Vensel, W.H., Hurkman, W.J., and McCormick, S. (2005) Proteome mapping of mature pollen of Arabidopsis thaliana. *Proteomics* 5, 4864-4884.
9. Li, J., Zhang, Z., Rosenzweig, J., Wang, Y.Y., & Chan, D.W. (2002) Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer. *Clinical chemistry* 48, 1296-1304.
10. Vemuri, P. K., Talluri, B., Sharma, A., Akkala, G., and Bodiga, V. L. (2015) Isolation and Characterization of a Lactose-Binding Lectin from Ocimum sanctum. *Journal of Applied Pharmaceutical Science* 5, 113-117.
11. Gupta, N., Martin, B.M., Metcalfe, D.D., & Rao, P.S. (1996) Identification of a novel hydroxyproline-rich glycoprotein as the major allergen in Partheniumpollen. *Journal of allergy and clinical immunology*, 98, 903-912.
12. Pan, S., Gu, S., Bradbury, E.M., and Chen, X. (2003) Single peptide-based protein identification in human proteome through MALDI-TOF MS coupled with amino acids coded mass tagging. *Analytical chemistry* 75, 1316-1324
13. Hellman, U., Wernstedt, C., Gonez, J., & Heldin, C. H. (1995) Improvement of an" In-Gel" digestion procedure for the micropreparation of internal protein fragments for amino acid sequencing. *Analytical biochemistry* 224, 451-455.
14. Mann, M., Ong, S. E., Gronborg, M., Steen, H., Jensen, O. N., and Pandey, A. (2002) Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome. *Trends in biotechnology* 20, 261-268.
15. Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research* 22, 4673-4680.
16. Kumar Vemuri., P and Veeravalli, S. (2014) Expression, Purification and Characterization of Human Recombinant Galectin 3 in Pichia pastoris. *Iranian Journal of Biotechnology* 12, 3-8.
17. Bufe, A., Schramm, G., Keown, M. B., Schlaak, M., and Becker, W. M. (1995) Major allergen Phl p Vb in timothy grass is a novel pollen RNase. *FEBS letters* 363, 6-12.
18. Ford, S. A., and Baldo, B. A. (1987) Identification of Bermuda grass (Cynodon dactylon)—pollen allergens by electroblotting. *Journal of allergy and clinical immunology* 79, 711-720.
19. Nakamura, R., and Teshima, R. (2013) Proteomics-based allergen analysis in plants. *Journal of proteomics* 93, 40-49.
20. Gangl, K., Niederberger, V., Valenta, R., and Nandy, A. (2015) Marker allergens and panallergens in tree and grass pollen allergy. *Allergo Journal International* 24, 158-169.
21. Campbell, B. C., Gilding, E. K., Timbrell, V., Guru, P., Loo, D., Zennaro, D., and Davies, J. M. (2015) Total transcriptome, proteome, and allergome of Johnson grass pollen, which is important for allergic rhinitis in subtropical regions. *Journal of Allergy and Clinical Immunology* 135, 133-142.