Available online at **www.joac.info**

ISSN: 2278-1862

Journal of Applicable Chemistry

2019, 8 (2): 511-516

(International Peer Reviewed Journal)

Mass Spectroscopy Based Proteomics for Identification of a Novel Alpha Amylase Inhibitor

K. S. N.Jyothi[1]* and Challa Suresh[2]

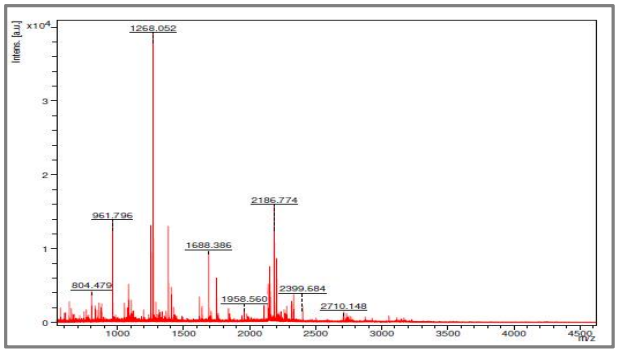
1. Department of Genetics and Biotechnology, UCW, Koti, Hyderabad, **INDIA**.
2. National Institute of Nutrition, Hyderabad, **INDIA**.

Email:ksn.jyothi@gmail.com, sureshnin2000@gmail.com

ABSTRACT

Mass Spectroscopy (MS) based Proteomics is a promising tool in macromolecular research. The wide range of protein activities, quantification of protein abundance, interactions and modifications, sequence analysis, comparative structural analysis can be probed effectively using MS. Most macromolecular research relies on the use of ESI or MALDI as an ionization source. The structural details or sequence information can be elucidated using Tandem MS. The utility of MS in protein and peptide analysis lies in its ability to provide highly accurate molecular weight information of intact molecules that is extremely useful for the characterization of newer proteins. A protein aceous alpha amylase inhibitor from a medicinal herb Oxalis corniculata has been isolated using standard protein purification techniques. The sequence analysis and structural determination was done using MS. The purified protein from the aqueous extract of Oxalis corniculata had a molecular weight of approximately 34Kda when separated on SDS PAGE which was further studied by MS analysis to get a peptide mass fingerprint (PMF). The peptide masses obtained were submitted to MASCOT search engine and the proteins similar to the putative alpha amylase inhibitor were observed. The isolated pure protein had 100% sequence similarity with PLDG3 protein of Arabidopsis thaliana. This MS based proteomic data paved way for the structural elucidation of the new protein. Further in silico analysis using Pair wise alignment and Virtual screening can help in building the structural and functional efficacy of the purified protein as an alpha amylase inhibitor.

Graphical Abstract



PMF of the purified protein

Keywords: Mass spectroscopy, Alpha amylase inhibitor, *Oxalis corniculata*, Virtual screening, Peptide Mass Fingerprint (PMF),

INTRODUCTION

Mass spectrometry (MS) is rapidly becoming a fundamental tool for biologists and biochemists in the protein identification and characterization. The utility of MS in protein and peptide analysis lies in its ability to provide highly accurate molecular weight information on intact molecules which is extremely useful for the characterization of newer proteins [**1, 2**].

The thrust area of Natural product research is identification and characterization of chemical constituents from plants that have high therapeutic value. One such class of compounds with potential demand is the alpha amylase inhibitors which are needed for the treatment of type II Diabetes mellitus. The present paper is an extension for a preliminary screening and purification study done on a small medicinal herb *Oxalis corniculata* [**3, 4**]. Mass spectrometry based proteomics was used to characterize the purified protein with proven enzymatic property. The use of mass spectrometry in protein characterization is indispensable. The accurate mass analysis of the constituent peptides obtained either by chemical or enzymatic treatment of the protein sample helps in unambiguous identification of the protein [**5**]. The two primary methods employed for ionization of whole proteins are Electrospray Ionization (ESI) and Matrix assisted laser desorption/ Ionization (MALDI).In ESI, intact proteins are ionized and then introduced to a mass analyzer. In MALDI, the proteins are enzymatically digested into smaller peptides using a protease such as Trypsin, which are then introduced into mass spectrometer and identified by Peptide mass fingerprinting or tandem mass spectrometry [**6, 7**]. The most widely used instrument for peptide mass analysis is the MALDI time-of- flight (TOF) instruments as they permit the acquisition of peptide mass fingerprints (PMFs) at high pace (1 PMF can be analyzed in approx. 10 sec).

Peptide mass fingerprinting (PMF): Also known as protein fingerprinting is an analytical technique for protein identification in which the unknown protein of interest is first cleaved into smaller peptides whose absolute masses can be accurately measured with a mass spectrometer such as MALDI-TOF or ESI-TOF. Peptide Mass fingerprinting uses the masses of proteolytic peptides as input to a search of a database of predicted masses that would arise from digestion of a list of known proteins [**8, 9**]. If a protein sequence in the reference list gives rise to a significant number of predicted masses that match the experimental values, there is some evidence that this protein was present in the original sample.

The peptide masses are compared to either a database containing known protein sequences or even the genome. This is achieved by using computer programs that translate the known genome of the organism into proteins, then theoretically cut the proteins into peptides, and calculate the absolute masses of the peptides from each protein [**10, 11**]. They then compare the masses of the peptides of the unknown protein to the theoretical peptide masses of each protein encoded in the genome. The results are statistically analyzed to find the best match [**12**]. The advantage of this method is that only the masses of the peptides have to be known and de novo peptide sequencing which is time consuming can be avoided. But if the protein sequence is not present in the database of interest then further *in-silico* analysis needs to be carried out to confirm the structure of the new protein.

MASCOT: It is a software search engine that uses mass spectrometry data to identify proteins from peptide sequence databases. Mascot identifies proteins by interpreting mass spectrometry data. A bottom-up approach is used where a protein sample is typically digested with Trypsin (a proteolytic enzyme) to form smaller peptides that can be measured by a mass spectrometer. Mascot compares the molecular weights of peptides that were measured by MS, against a database of known peptides [**13, 14**]. The program cleaves every protein in the specified search database *insilico* according to specific rules depending on the cleavage enzyme used for digestion and calculates the theoretical mass for each peptide. Mascot then computes a score based on the probability that the peptides from a sample match those in the selected protein database. The more peptides Mascot identifies from a particular protein, the higher the Mascot score for that protein. Based on the Mascot search analysis, the sequence similarity studies can be done which can further be used for structure prediction of the

newer protein. Structural studies of proteins have gone a long way with the utilization of Mass spectrometry based proteomic analysis [**15, 16**].

MATERIALS AND METHODS

Instrumentation: The peptide mass fingerprinting analysis was done at the Sandor Proteomics lab, CDFD, Hyderabad. The instrument used was MALDI-TOF/TOF MS Bruker Daltonics, ULTRAFLEX III.

Preliminary analysis: To avoid complications in the analysis, an isolated protein obtained from SDS-PAGE was utilized for the PMF based protein identification. The preliminary work done, was screening of many indigenous medicinal plants for the identification of an efficient alpha amylase inhibitor from indigenous medicinal plants. Standard protein purification techniques were performed to isolate a purified proteinaceous alpha amylase inhibitor from a small medicinal herb *Oxalis corniculata* (Jyothi *et al*). The purified protein was then separated on a SDS-PAGE. The protein band obtained from SDS-PAGE at about 34KDa has been selected for analysis by MALDI-TOF.

Pre-Treatment of the sample: The sample obtained from the SDS gel was trypsin digested and the peptides obtained were mixed with HCCA matrix in 1:1 ratio and the resulting 2ul was spotted onto the MALDI plate. After air drying the sample, it was analyzed on the MALDI TOF/TOF ULTRAFLEX III instrument with a laser pulse duration of 1-5ns, at a laser attenuator range 20%, ion source voltage 25kv and laser frequency 100 Hz. Based on the m/z ratios obtained, further analysis was done with FLEX ANALYSIS SOFTWARE for obtaining the PMF. The masses obtained in the peptide mass fingerprint were submitted for **Mascot** search in "concerned" database for identification of the protein as part of the *De novo* peptide sequencing.

RESULTS AND DISCUSSION

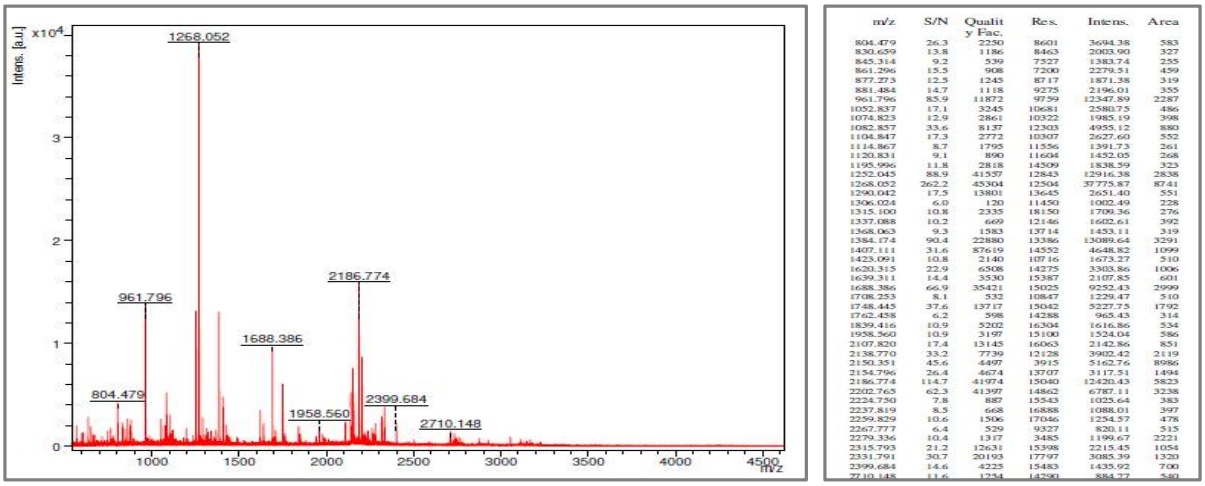
The purified protein obtained on SDS-PAGE at a molecular weight of approximately 34 Kda was studied for its Peptide Mass fingerprint (PMF) by MSMS analysis. The peptide masses obtained were submitted to Mascot search engine and the proteins similar to the putative AI-1 were given as the result. The sequence of amino acids of the peptide fragments produced by the trypsin digest is given in figure 1. The spectrum as obtained from the MSMS analysis and the m/z ratios of the peptide fragments obtained from the trypsin digestion are given in the figure 2.

The mass spectrometric analysis produced a list of molecular weights of the fragments which is often called a peak list. The peptide masses were compared to protein database Mascot, which contain protein sequence information. Software performed *in silico* digests on proteins in the database with the same enzyme (e.g. trypsin) used in the chemical cleavage reaction. The mass of these peptide fragments is then calculated and compared to the peak list of measured peptide masses. The results are statistically analyzed and possible matches are returned in a results table (Figure 3). The protein analyzed by MS in the present work is isolated and purified from a medicinal herb *Oxalis corniculata*. To ascertain its role as an alpha amylase inhibitor, sequence determination and structural studies were needed which was possible using Mass Spectrometry. MS based approaches are now the best method of choice for protein identification because of the improvements in in genomics and protein databases. Proteomic analysis is assisted seamlessly by excellent MS data processing algorithms.

The Mascot results given for the sequence showed its 100% similarity with PLDG3 protein of *Arabidopsis thaliana*. To check the precise similarity of the sequence of interest with that of the known alpha amylase inhibitors from the protein database, pairwise alignment was performed. The similarity search with pair wise alignment followed by secondary structure prediction helped in designing a suitable secondary structure for the given protein of interest. Further *in-silico* docking studies helped in establishing the alpha amylase inhibitory role of the purified protein.

Query	Start - End	Observed	Mr(expt)	Mr(calc)	ppm	M	Score	Peptide
51	1 - 27	3053.4594	3052.4521	3051.2837	383	0		-.MAYHPVYNETMSMGGGSSNEFGQWLDK.Q + Oxidat
23	28 - 40	1407.1108	1406.1035	1405.7202	273	0	13	K.QLVPFDTSSGSLR.V
33	182 - 199	2107.8203	2106.8130	2107.0918	-132	1		K.QGAVLSLSIQYIPMERMR.L + Oxidation (M)
34	204 - 223	2138.7703	2137.7630	2138.0983	-157	1		K.GVGFGVECVGVPGTYFPLRK.G
24	256 - 266	1423.0913	1422.0841	1421.6180	328	1		K.CWEDMADAIRR.A
27	270 - 283	1688.3857	1687.3784	1686.9246	269	0	8	R.LIYITGWSVFHPVR.L
15	330 - 340	1252.0446	1251.0373	1251.5401	-402	0		K.GLMNTSDEETR.R
16	330 - 340	1268.0519	1267.0446	1267.5350	-387	0	10	K.GLMNTSDEETR.R + Oxidation (M)
39	450 - 468	2224.7501	2223.7428	2224.0623	-144	1		K.IDGPAAYDVLANFEERWMK.A
48	609 - 632	2732.1627	2731.1554	2730.3839	283	1		R.EKFAAYIVIPMWPEGAPTSNPIQR.I + Oxidation
29	695 - 710	1748.4446	1747.4374	1748.0057	-325	1	13	R.KPPQLNAAQVQALKSR.R

Figure 1. The sequence of amino acids obtained from the peptide fragments.



m/z	S/N	Qualit y Fac.	Res.	Intens.	Area
804.479	26.3	2250	8601	3694.38	583
830.639	13.8	1186	8463	2003.90	327
845.314	9.2	539	7527	1383.74	255
861.296	15.5	908	7200	2279.51	459
877.273	12.3	1245	8717	1871.38	319
881.484	14.7	1118	9275	2196.01	355
961.796	85.9	11872	9759	12347.89	2287
1052.837	17.1	3245	10681	2580.75	486
1074.823	12.9	2861	10322	1985.19	398
1082.857	33.6	8137	12303	4955.12	880
1104.847	17.3	2772	10307	2627.60	552
1114.867	8.7	1795	11556	1391.73	261
1120.831	9.1	890	11604	1452.05	268
1195.996	11.8	2818	14509	1838.59	323
1252.045	88.9	41557	12843	12916.38	2838
1268.052	262.2	45304	12504	37775.87	8741
1290.042	17.5	13801	13645	2651.40	551
1306.024	6.0	120	11450	1002.49	228
1315.100	10.8	2335	18150	1709.36	276
1337.088	10.2	669	12146	1602.61	392
1368.063	9.3	1583	13714	1453.11	319
1384.174	90.4	22880	13386	13089.64	3291
1407.111	31.6	87619	14552	4648.82	1099
1423.091	10.8	2140	10716	1673.27	510
1620.315	22.9	6508	14275	3303.86	1006
1639.311	14.4	3530	15387	2107.85	601
1688.386	66.9	35421	15025	9252.43	2999
1708.253	8.1	532	10847	1229.47	510
1748.445	37.6	13717	15042	5227.75	1792
1762.458	6.2	598	14288	965.43	314
1839.416	10.9	5202	16304	1616.86	534
1958.560	10.9	3197	15100	1524.04	586
2107.820	17.4	13145	16063	2142.86	851
2138.770	33.2	7739	12128	3902.42	2119
2150.351	45.6	4497	3915	5162.76	8986
2154.796	26.4	4674	13707	3117.51	1494
2186.774	114.7	41974	15040	12420.43	5823
2202.765	62.3	41397	14862	6787.11	3238
2224.750	7.8	887	15543	1025.64	383
2237.819	8.5	668	16888	1088.01	397
2259.829	10.6	1506	17046	1254.57	478
2267.777	6.4	529	9327	820.11	515
2279.336	10.4	1317	3485	1199.67	2221
2315.793	21.2	12631	15398	2215.45	1054
2331.791	30.7	20193	17797	3085.39	1320
2399.684	14.6	4225	15483	1435.92	700
2710.148	11.6	1254	14790	884.27	540

Figure 2. PMF of the purified protein and m/z ratios of the peptide fragments.

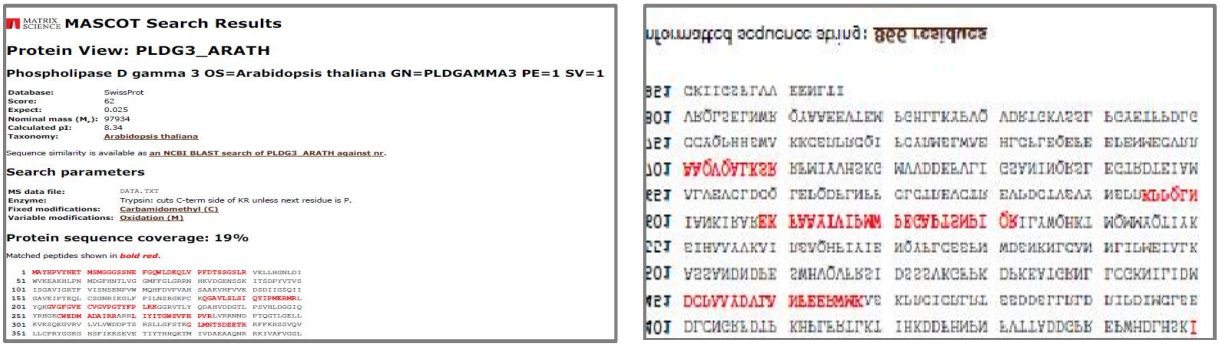


Figure 3. MASCOT Search Results.

APPLICATION

Mass spectrometry is a widely used technique for protein identification, de novo protein sequencing, protein quantitation and protein structure prediction. Given the complexity of the protein mixtures that coexist in biological medium, their structural determination is a major challenge [17, 18]. Mass spectroscopy has evolved as a very promising tool in overcoming many obstacles in protein structural studies provided the researcher gets a pure protein sample either by two dimensional gel electrophoresis or HPLC [19]. In the present work, Mass spectrometric analysis and the PMF helped in further analysis of the alpha amylase inhibitory role of the purified protein. Without MS it would be

very difficult to obtain an overall picture of the proteome [20-22]. Mass spectroscopy based proteome research has gone a long way in proving itself as a central life science technology in identification, quantification, and characterization of the proteins that constitute a proteome.

CONCLUSION

Analytical advances in Mass Spectrometry largely contributed to the life science, medical and pharmaceutical fields. It contributed to the remarkable development of antibody drugs and their therapeutic applications. Proteomic analysis is greatly assisted by excellent MS data processing algorithms which largely helped in the development of clinical biomarkers and natural product discovery. Many such plant-based natural products like alpha amylase inhibitors from medicinal plants are of great value in the treatment of Type II Diabetes mellitus. Given the contraindications of the anti-diabetic drugs available in the market, there is extensive research going on in the field of natural product research for identification of natural plant based anti-diabetic drugs, an important class of which are alpha amylase inhibitors. Many plant based amylase inhibitors are proteins in nature. Their characterization and structure prediction largely relies on the modern Mass spectrometry and its associated techniques. The advancements in the field of biological research can be greatly attributed to the development of precise techniques like mass spectrometry based proteomics, metabolomics and systems biology. Integrated biology approaches and intervention of complex networking analysis in biological systems along with modern spectroscopic techniques, like Mass spectrometry, made research in life sciences taken a new direction and in bringing out significant progress in understanding complex protein molecules. Thus, MS based techniques provide a versatile platform for the study of protein structure and dynamics. A greater wealth of information associated with near complete proteome coverage in humans and many other organisms is mainly because of mass spectroscopy based proteome research stored in databases and bioinformatics resources which is of immense help for research in varied fields of science.

REFERENCES

[1]. K. S. N Jyothi, M. Shailaja, J. Viveni, C. Suresh. Identification of a Proteinaceous Alpha Amylase Inhibitor from a Medicinal Herb Oxalis corniculata L. (Oxalidaceae). *J. Homeop Ayurv. Med.*, **2014**, 3, 165-170.

[2]. K. S. N Jyothi, P. Hemalatha, S. Challa. Evaluation of α-amylase inhibitory potential of three medicinally important traditional wild food plants in India, *Int. J. green pharmacy*, **2011**, 95-99.

[3]. W. S. Noble, M. J. MacCoss, Computational and Statistical Analysis of Protein Mass Spectrometry Data, *PLoS Comput Biol.*, **2012**, 8, 1.

[4]. T. S. Price, M. B. Lucitt, W. Wu, D. J. Austin, A. Pizarro, EBP, a program for protein identification using multiple tandem mass spectrometry datasets, *Mol Cell Proteomics*, **2007**, 6, 527–536.

[5]. A. I. Nezvizhskii, A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics, *Journal of Proteomics*, **2010**, 73, 2092–2123.

[6]. A. Leitner, Crosslinking and mass spectrometry: an integrated technology to understand the structure and function of molecular machines, *Trends Biochem. Sci.* **2016**, 41, 20–32.

[7]. A. Politis. et al. (2014) A mass spectrometry-based hybrid method for structural modeling of protein complexes, *Nat. Methods*, **2014**, 403–406.

[8]. J. Pablo, Villacorta, Antonio Salmerón-García, David A. Pelta, José Cabeza, Antonio Lario and Natalia Navas, Cluster-based comparison of the peptide mass fingerprint obtained by MALDI-TOF mass spectrometry. A case study: long-term stability of rituximab, *Analyst*, **2015**,140, 1717-1730.

[9]. A. Sinz, Chemical cross-linking and native mass spectrometry: a fruitful combination for structural biology. *Protein Sci.*, **2015**, 24, 1193–1209

[10]. A. H. Smits, M. Vermeulen, Characterizing protein– protein interactions using mass spectrometry: challenges and opportunities, *Trends Biotechnol*., **2016**, 34, 825–834
[11]. S. Ovchinnikov, Protein structure determination using metagenome sequence data, *Science.* **2017**, 355, 294–298
[12]. R. Aebersold, M. Mann. Mass-spectrometric exploration of proteome structure and function, *Nature*, **2016**, 537, 347-355.
[13]. J. D. Lapek Jr., P. Greninger, R. Morris, A.Amzallag, I. PruteanuMalinici, C. H. Benes, W. Haas, Detection of dysregulated protein-association networks by high-throughput proteomics predicts cancer vulnerabilities, *Nature Biotechnology*, **2017**, 35, 983-989.
[14]. P. Giansanti, L. Tsiatsiani, T. Y. Low, A. J. Heck, Six alternative proteases for mass spectrometry-based proteomics beyond trypsin, *Nature Protocols*, **2016**, 11, 993-1006.
[15]. L. Sun, A. S. Hebert, X. Yan, Y. Zhao, M. S. Westphall, M. J. Rush, N. J. Dovichi, Over 10,000 peptide identifications from the HeLa proteome by using single-shot capillary zone electrophoresis combined with tandem mass spectrometry, *Angewandte Chemie*, **2014**, 53, 13931-13933
[16]. F. Calderon-Celis, J. R. Encinar, A. Sanz-Medel, Standardization approaches in absolute quantitative proteomics with mass spectrometry, *Mass Spectrometry Reviews*, **2017**, 1-23.
[17]. P. Hommerson, A. M. Khan, G. J. de Jong, G. W. Somsen, Ionization techniques in capillary electrophoresis-mass spectrometry: Principles, design, and application, *Mass Spectrometry Reviews*, **2011**, 30, 1096-1120
[18]. M. Iwasaki, S. Miwa, T. Ikegami, M. Tomita, N. Tanaka, Y. Ishihama, One-dimensional capillary liquid chromatographic separation coupled with tandem mass spectrometry unveils the *Escherichia coli* proteome on a microarray scale, *Analytical Chemistry*, **2010**, 82, 2616-2620.
[19]. A. Guthals, N. Bandeira, Peptide identification by tandem mass spectrometry with alternate fragmentation modes, *Molecular and Cellular Proteomics*, **2012**, 11, 550-557.
[20]. C. E. Bakalarski, D. S. Kirkpatrick, A biologist's field guide to multiplexed quantitative proteomics, *Molecular and Cellular Proteomics*, **2016**, 15, 1489-1497.
[21]. E. Caron, D. J. Kowalewski, C. Chiek Koh, T. Sturm, H. Schuster, R. Aebersold, Analysis of Major Histocompatibility Complex (MHC) immunopeptidomes using mass spectrometry, *Molecular and Cellular Proteomics*, **2015**, 14, 3105-3117.
[22]. M. Matsumoto, F. Matsuzaki, K. Oshikawa, N. Goshima, M. Mori, Y. Kawamura, K. I. Nakayama, A large-scale targeted proteomics assay resource based on an in vitro human proteome, *Nature Methods*, **2017**, 14, 251-258.