THE PROTEOMICS REVOLUTION: ADVANCES AND APPLICATIONS

Syed Sara Afreen, *, Aarthi Saima Ghanta, Swarjan Paul Ummadisetti, Jagdish Chettigari, Durga Pani Kumar Anumolu , Ashok Gorja.

Department of Pharmaceutical Analysis, Gokaraju Rangaraju College of Pharmacy, Osmania University, Hyderabad, Telangana- 500090.

***Corresponding author**

Syed Sara Afreen

Asst Professor, Department of Pharmaceutical Analysis

Gokaraju Rangaraju College of Pharmacy,

Hyderabad ,Telangana-500090, India

ABSTRACT

The extensive study of proteins, or proteomics, has become a vital field for comprehending disease mechanisms, cellular processes, and the creation of new treatments. Proteins in complicated biological samples can now be identified, quantified, and characterised thanks to recent developments in high-throughput techniques, especially mass spectrometry and protein microarrays. Protein expression, post-translational changes, protein-protein interactions, and cellular localisation may all be explored thanks to this thorough study, which offers hitherto unachievable levels of detail on cellular processes.The complexity of the proteome, the variety of protein isoforms, and the dynamic character of protein interactions are some of the obstacles that still exist despite tremendous advancements. Furthermore, proteomics' integration with other "omics" technologies—such as transcriptomics, metabolomics, and genomics—is still in its infancy, with computational tools for data.

INTRODUCTION

Protein, highly complex substance that is present in all living organisms. Proteins are the polymers of amino acids. Emil Fischer and Franz Hofmeister, reported about proteins in 1902, Proteins play an important role in metabolic activities. Primary structure of protein is determined by the sequence of specific amino acids, encoded by the mRNA, which directs the proper folding of the polypeptide chain into the secondary structure. One type of secondary structure is the alpha helix, a region of the polypeptide that folds into a corkscrew shape. Beta strands are linear structures of polypeptides, bonding together to form a flat beta sheet. Turns and coils interact chemically with each other to form the unique three dimensional shape of the proper three-dimensional structure creates the final protein. Many proteins, however, have several different polypeptide subunits that make the final active protein. For these proteins, the interactions between the different subunits form the quaternary structure. One of the most promising developments to come from the study of human genes and proteins has been the identification of potential new drugs for thetreatments of disease.

This relies on genome and proteome information to identify proteinsassociated with a disease. The term “proteomics” was first coined in 1995 and was defined as the large-scale characterization of the entire protein complement of a cell line, tissue, or organism[1,2,3]. Proteomics is the large-scale study of proteins particularly their composition, structures,

functions, and interactions of the proteins directing the activities of cell[4,5].The main theme of interest proteomics it gives a much better understanding of an organism thangenomics. Genomics can give a rough estimation of expression of a protein. Most of the proteins function in collaboration with other proteins, and the main goal of proteomics are to identify which proteins interact. After genomics, proteomics is often considered as the advanced step in the study of biological systems. It is much more complicated than genomics, mostly because while an organism's genome is more or less constant, the total protein expression profile always changes with time, micro and macro environmental conditions. The dynamic role of molecules to support the life is documented since the initial stages of biological research. To demonstrate the importance of these molecules, Berzelius in 1838 given the title "protein", which is originated from the Greek word, proteios, meaning "the first rank"[6]. The "proteome" can be defined as the overall protein content of a cell that is characterized with regard to their localization, interactions, post-translational modifications and turnover, at a particular time. The term "proteomics" was first used by Marc Wilkins in 1996 to denote the "Protein in complement of a genome"[7] .

Proteomics is crucial for early disease diagnosis, prognosis and to monitor the disease development. Furthermore, it also has a vital role in drug development as target molecules. Proteomics is the characterization of proteome, including expression, structure, functions, interactions and modifications of proteins at any stage[8]. The proteome also fluctuates from time to time, cell to cell and in response to external stimuli. Proteomics in eukaryotic cells is complex due to post-translational modifications, which arise at different sites by numerous ways[9].Proteomics can analyze the expression of a protein at different levels allowing theassessment of specific quantitative and qualitative cellular responses related to that protein[10]. Qualitative and quantitative proteomes are measured at post-transcriptional, transcriptomic, and genomic levels[11].Proteomics offers a superior understanding of an organism's structure and function compared to genomics, despite its greater complexity, as protein expression undergoes changes over time and in response to environmental conditions [12]. Relying solely on the study of genes makes it impossible to acquire various types of information. For example, elucidating the mechanisms behind disease development, aging, and the impacts of environmental factors is not achievable solely through genome studies. Moreover, the identification of drugtargets and the characterization of protein modifications are possible only through the examination of proteins[13].

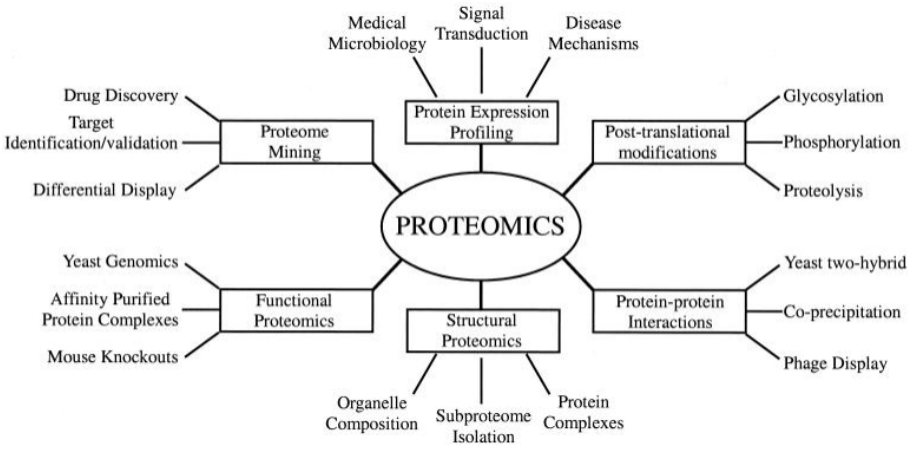

Figure 1.1- Featuring of Proteomics.

2. TYPES OF PROTEOMICS

Based on the protein response under stress conditions proteomics are classified into different groups.

2. 1. Expression proteomics

Expression proteomics is used to study the qualitative and quantitative expression of total proteins under two different conditions. Like the normal cell and treated or diseased cell can be compared to understand the protein that is responsible for the stress or diseased state or the protein that is expressed due to disease. Typically, expression proteomics studies are addressed to the investigation of the expression protein patterns in abnormal cells. Ex. Compare tumor tissue sample and the normal tissue can be analyzed for differential protein expression. 2-D gel electrophoresis, mass spectrometry technique were used to observed the protein expressional changes, which is present and absent in tumor tissue, when compared with normal tissue. Which are over expressed and under expressed can be identified and characterized protein activities multi-protein complexes, and signalling pathways[14,15]. Identification of these proteins will give valuable information about molecular biology of tumor formation and disease-specific manner for use as diagnostic markers or therapeutic targets.

2. 2. Structural proteomics

Structural proteomics helps to understand three-dimensional shape and structural complexities of functional proteins. Structural prediction of a protein when its amino acid

sequence is determined directly by sequencing or from the gene with a method called homology modelling. Structural proteomics can give detailed information about the structure and function of protein complexes present in a specific cellular organelle. It is possible to identify all the proteins present in a complex system such as membranes, ribosomes, and cell organelles and to characterise all the protein interactions that can be possible between these proteins and protein complexes. Different technologies such as X-ray crystallography and NMR spectroscopy were mainly used for structure determination[16].

2. 3. Functional proteomics

Functional proteomics explains understanding the protein functions as well as unrevealing molecular mechanisms within the cell then depend on the identification of the interacting protein partners. The association of an unknown protein with partners belonging to a specific protein complex involved in a particular mechanism would in fact, be strongly suggestive of its biological function[17,18]. Furthermore, detailed description of the cellular signalling pathways might greatly benefit from the elucidation of protein- protein interactions in-vivo.

2. 4. Techniques involved in proteomics

In proteomic analysis both analytical and bio-informatics tools were used to characterize protein structure and functions. Analytical techniques 2-D gel electrophoresis, MALDI-TOF-MS were used. In case of bio-informatics numbers of software tools were used.

2. 5. 2-D gel electrophoresis

In 2-D gel electrophoresis, protein samples are resolved based on charge, in a step called isoelectric focusing, and then based on molecular weight in second step[19]. The result is an image in thousands of small spots, each representing a protein. A good 2-D gel can resolve one thousand to two thousand protein spots, which appear after staining, as dots in the gel. 2-D gel electrophoresis technique is mainly used to compare two similar samples to find specific protein differences.

2. 6. MS analysis

Mass spectrometry is an analytical technique that produces spectra of the masses of the atoms or molecules comprising a sample of material. The spectra are used to determine the elemental or isotopic signature of a sample, the masses of particles and of molecules, and to elucidate the chemical structures of molecules, such as peptides and other chemical

compounds. Mass spectrometry works by ionizing chemical compounds to generate charged molecules or molecule fragments and measuring their mass to charge ratios. MALDI-TOF is the most useful technique for protein identification.

2. 7. MALDI-TOF-MS

Matrix Assisted Laser Desorption/Ionisation is a soft ionization technique used in spectrometry, allowing to analysis the biomolecules like DNA, protein, peptides. Biomolecules and synthetic polymers have low volatility and are thermally unstable, which has limited the use of MS as a means of characterization. These problems have been minimized through the development of MALDI-TOF MS, which allows for the massdetermination of biomolecules by ionization and vaporization without degradation, a Laser beam used to ionize the sample[20].

High-throughput methods:

1. **Analytical, functional and reverse-phase microarrays**

Protein microarrays apply small amounts of sample to a “chip” for analysis (this is sometimes in the form of a glass slide with a chemically modified surface). Specific antibodies can be immobilized to the chip surface and used to capture target proteins in a complex sample. This is termed an analytical protein microarray, and these types of microarrays are used to measure the expression levels and binding affinities of proteins in a sample. Functional protein microarrays are used to characterize protein functions such as protein–RNA interactions and enzyme-substrate turnover. In a reverse-phase protein microarray, proteins from e.g., healthy vs. diseased tissues or untreated vs. treated cells are bound to the chip, and the chip is then probed with antibodies against the target proteins.

2. Mass spectrometry-based proteomics

There are several “gel-free” methods for separating proteins, including isotope-coded affinity tag (ICAT), stable isotope labelling with amino acids in cell culture (SILAC) and isobaric tags for relative and absolute quantitation (iTRAQ). These approaches allow for both quantitation and comparative/differential proteomics. There are also other, less quantitative techniques such as multidimensional protein identification technology (MudPIT), which offer the advantages of being faster and simpler. Other gel-free, chromatographic techniques for protein separation includegas chromatography(GC) and liquid chromatography (LC).

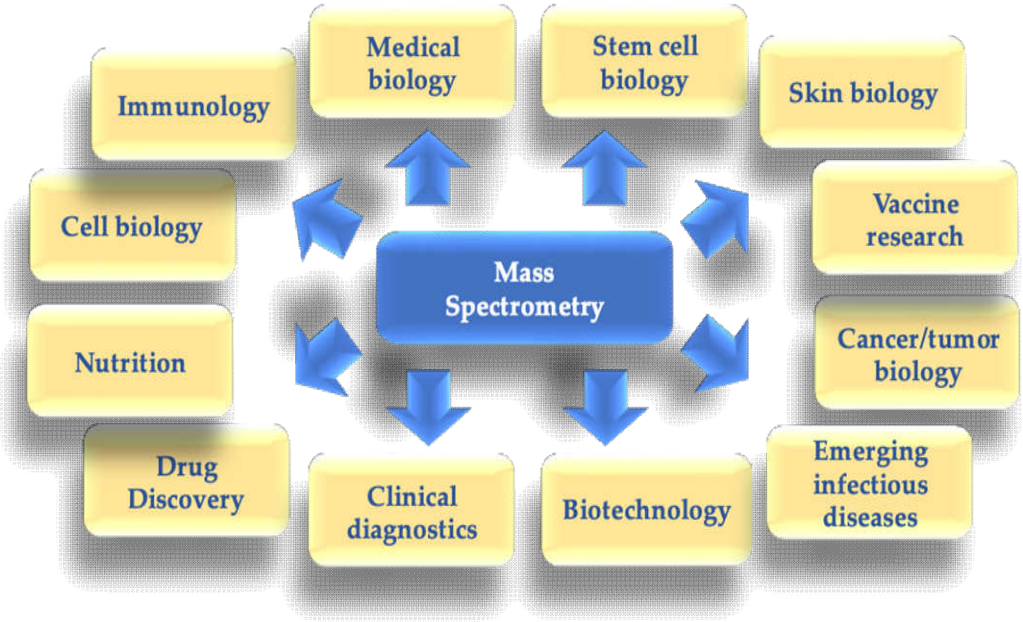

Figure 1.2 - Mass spectroscopy

3. ADVANCED METHODS IN PROTEOMICS

3.1 Isotope-coded affinity tags (ICAT)

It is a gel- free method for quantitative proteomics that relies on chemical labelling reagents. These chemical probes consist of three general elements i.e. defined amino acid side chain, an isotopically coded linker, and a tag for the affinity isolation of labelled proteins/peptides. For quantitative comparison of two proteomes, sample labelled with isotopically light, and other one is heavy version. Both samples combined with isotope-coded tagging reagents. These peptides are analyzed by LC-MS. Tags were deuterium, 13C 18. The technique mainly used the relative quantification of proteins present in two or more biological samples. Visible isotope-coded affinity tags are the additional method in ICATVisible tag that allows the electrophoresis position of tagged peptides to be easily monitored.

3.2 Isobaric Tags for Relative and Absolute Quantification (iTRAQ)

Isobaric tags for relative and absolute quantitation (iTRAQ), it is also a non- gel- based technique used to quantify proteins. iTRAQ is used in proteomics to study quantitative changes in the proteome 19. Based on the covalent labelling of the N-terminus and side chain amines of peptides from protein digestions with tags of varying mass, 4-plex and 8-plex are the reagents can be used to label all peptides from different samples. The samples can be

analysed by using mass spectrometry MS/MS. Different types of software's are available for analyse the MS/MS spectra's i.e. j-Tracker, j-TraqX[21].

3.3 Absolute Quantification (AQUA)

AQUA, studies the absolute quantification of proteins and their modification sates. Covalent modifications can be used to prepare synthetic proteins. These modifications are chemically identical to naturally occurring posttranslational modifications. These types of peptides used to quantify the post translational modified proteins after proteolysis with the help of tandem mass spectrometer.

3.4 ESI-Q-IT-MS

Micro electrospray ionization (ESI)-Quadrupole ion trap (QIT) Time of flight (TOF) mass spectrometer (MS) has a very good resolution. In ESI ionisation proteins are ionised in solution and carry multiple charge state. The advantage of using ESI-QTOF analysis for protein mass determination is that due to the high charge state of proteins their m/z measurements is typically less than 2000 and the TOF detector has a very good mass accuracy in this scan range. This result is more accurate mass measurements for proteins in ESI-QTOF.

3.5 SELDI-TOF-MS

The technique Surface-enhanced laser desorption/ionization (SELDI) is used for the analysis of protein mixtures, it is an ionization method in mass spectrometry21. SELDI is typically used with time-of-flight mass spectrometers and is used to detect proteins in clinical samples; to compare protein levels with and without a disease can be used for biomarker discovery.

Sample preparation for proteomics

Preparation of sample is the most fundamental step in proteomics research that considerably affects the results of an experiment. Therefore, the selection of appropriate experimental model and sample preparation method is essential for reliable results, especially in comparative proteomics, that deal with the minor variances of experimental samples compared with the control. The major impediments associated with the analysis of complex biological materials are the wider range of protein abundance. A particular cell could have only few copies of a protein, but we may expect up to million copies of an abundant protein

therefore these abundant proteins should be removed for most of the proteomic analysis. The Pre analytical samples treatment include various methods for fractionation and proteins enrichment could be helpful in this regard.

3. 6. Applications of proteomics

Proteomics is widely used technique in biological fields, mainly applied in Oncology (Tumour biology), Biomedicine, Agriculture and Food Microbiology.

3. 7. Oncology

Oncology refers study of Tumor cell, Tumor metastasis, is the process spread of cancer from one organ to another non-adjacent organ cause death in patients[22]. The major challenge in medicine to describe the molecular and cellular mechanisms underlying tumor metastasis. Analyse the protein expressions correlated to the metastatic process which help to understand the mechanism of metastasis and thus facilitate the development of strategies for the therapeutic interventions and clinical management of cancer. Proteomics is systematic research, the main aim of this research is to characterize the protein expressions, functions of tumor cells and widely used in biomarker discovery.

3. 8. Bio-medical applications

The study of interactions between microbial pathogens and their hosts is called “infectomics”. It is very interesting area in proteomics. It deals with the fundamentals of the infection’s origin and their effect on organs. The main aim of this research is to prevent or cure disease at starting level. Advanced diagnostic issues related to emerging infections, increasing of fastidious bacteria, and generation of patient- tailored phenotypes[22].

3. 9. Agricultural applications

The applications of plant proteomics scientific research is still in budding stage. Proteomics is also used to know plant-insect interactions that help identify candidate genes involved in the defensive response of plants to herbivore[23]. Population growth and effect of global climate changes imposing severe limits on the sustainability of agricultural crop production.

3. 10. Food Microbiology

The use of proteomics in food technology is presented especially for characterisation and standardisation of raw materials, process development, and detection of batch-to batch variations and quality control of the final product. Further attention is paid to the aspects of food safety, especially regarding biological and microbial safety and the use of genetically modified foods[24].

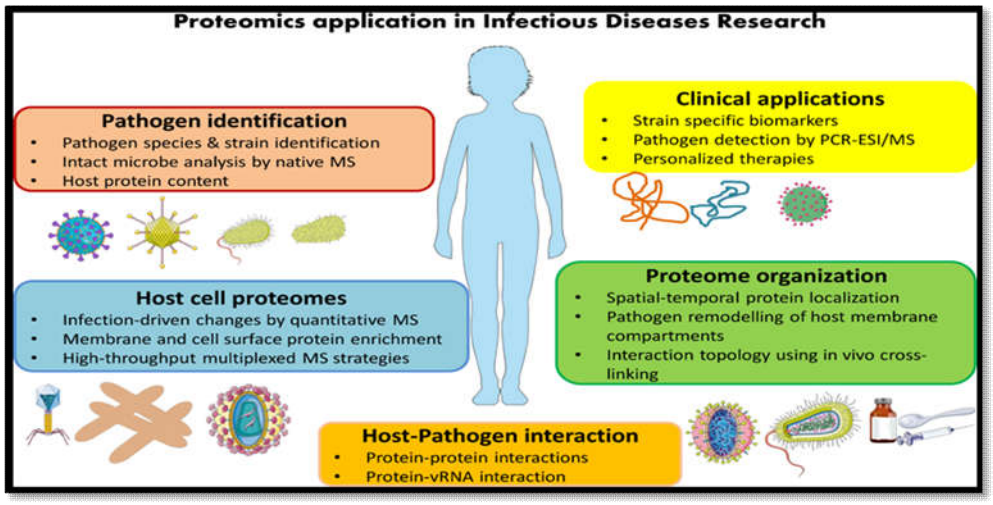

Figure 1.3- Different applications and usage of proteomics in medical diagnosis.

CONCLUSION :

Based on the results, the current review came to the conclusion that proteomics applications are applicable to every biological activity and offer a way to make better use of the produced protein data.Proteomics has significantly improved our knowledge of disease mechanisms, biological processes, and treatment approaches. Complex protein networks in a range of biological situations may now be identified and quantified thanks to the use of high-throughput technologies like mass spectrometry and protein microarrays. Proteomic investigations are now much more accurate and efficient thanks to these technological developments and better computational tools for data analysis.

ACKNOWLEDGEMENT:

The authors are thankful to the management and Principal Dr.M.Ganga Raju of Gokaraju Rangaraju College of Pharmacy for providing facilities for this review work.

REFERENCES

[1] Anderson NG, Anderson NL. Twenty years of two-dimensional electrophoresis: past, present and future. Electrophoresis. 1996 Mar 1; 17(3):443-53.

[2] Wasinger VC, Cordwell SJ, Cerpa-Poljak A, Yan JX, Gooley AA, Wilkins MR, Duncan MW, Harris R, Williams KL, Humphery-Smith I. Progress with gene-product mapping of the Mollicutes: Mycoplasma genitalium. Electrophoresis. 1995;16(1):1090-4.

[3] Wilkins MR, Sanchez JC, Gooley AA, Appel RD, Humphery-Smith I, Hochstrasser DF, Williams KL. Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. Biotechnology and genetic engineering reviews. 1996 Dec 1;13(1):19-50.

[4] Anderson NL, Anderson NG. Proteome and proteomics: new technologies, new concepts, and new words. Electrophoresis. 1998 Aug;19(11):1853-61.

[5] Blackstock WP, Weir MP. Proteomics: quantitative and physical mapping of cellular proteins. Trends in biotechnology. 1999 Mar 1;17(3):121-7.

[6] Cristea IM, Gaskell SJ, Whetton AD. Proteomics techniques and their application to hematology. Blood. 2004 May 15;103(10):3624-34.

[7] Wilkins MR, Sanchez JC, Gooley AA, Appel RD, Humphery-Smith I, Hochstrasser DF, Williams KL. Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. Biotechnology and genetic engineering reviews. 1996 Dec 1;13(1):19-50.

[8] Alhendi AS. A Review: Protein Identification by LC-MS: Principles, Instrumentation, and Applications. Iraqi Journal of Science. 2020 Oct 1;61(10).

[9] Aslam B, Basit M, Nisar MA, Khurshid M, Rasool MH. Proteomics: technologies and their applications. Journal of chromatographic science. 2016 Oct 18:1-5.

[10] Xiao GG, Recker RR, Deng HW. Recent advances in proteomics and cancer biomarker discovery. Clinical medicine. Oncology. 2008 Jan;2:CMO-S539.

[11] Low TY, van Heesch S, van den Toorn H, Giansanti P, Cristobal A, Toonen P, Schafer S, Hübner N, van Breukelen B, Mohammed S, Cuppen E. Quantitative and qualitative proteome characteristics extracted from in-depth integrated genomics and proteomics analysis. Cell reports. 2013 Dec 12;5(5):1469-78.

[12] Holman JD, Dasari S, Tabb DL. Informatics of protein and posttranslational modification detection via shotgun proteomics. Proteomics for Biomarker Discovery. 2013:167-79.

[13] Graves PR, Haystead TA. Molecular biologist's guide to proteomics. Microbiology and molecular biology reviews. 2002 Mar;66(1):39-63.

[14] Pandey A, Fernandez MM, Steen H, Blagoev B, Nielsen MM, Roche S, Mann M, Lodish HF. Identification of a novel immunoreceptor tyrosine-based activation motif-containing molecule, STAM2, by mass spectrometry and its involvement in growth factor and cytokine receptor signalling pathways. Journal of Biological Chemistry. 2000 Dec 8;275(49):38633-9.

[15] Hinsby AM, Olsen JV, Bennett KL, Mann M. Signalling initiated by overexpression of the fibroblast growth factor receptor-1 investigated by mass spectrometry. Molecular & Cellular Proteomics. 2003 Jan 1;2(1):29-36.

[16] Yokoyama S, Hirota H, Kigawa T, Yabuki T, Shirouzu M, Terada T, Ito Y, Matsuo Y, Kuroda Y, Nishimura Y, Kyogoku Y. Structural genomics projects in Japan. nature structural biology. 2000 Nov;7(11):943-5.

[17] Gavin AC, Bösche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick JM, Michon AM, Cruciat CM, Remor M. Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature. 2002 Jan 10;415(6868):141-7.

[18] Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, Adams SL, Millar A, Taylor P, Bennett K, Boutilier K, Yang L. Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. Nature. 2002 Jan 10;415(6868):180-3.

[19] Chandrasekhar K, Dileep A, Lebonah DE, Pramoda Kumari J. A short review on proteomics and its applications. International Letters of Natural Sciences. 2014;12(1).

[20] Wu KJ, Shaler TA, Becker CH. Time-of-flight mass spectrometry of underivatized single-stranded DNA oligomers by matrix-assisted laser desorption. Analytical chemistry. 1994 May 1;66(10):1637-45.

[21] Muth T, Keller D, Puetz SM, Martens L, Sickmann A, Boehm AM. jTraqX: a free, platform independent tool for isobaric tag quantitation at the protein level. Proteomics. 2010 Mar;10(6):1223-5.

[22] Chiang AC, Massagué J. Molecular basis of metastasis. New England Journal of Medicine. 2008 Dec 25;359(26):2814-23.

[23] Alagarsamy M, Amal TC, Karuppan S, Adhimoolam K. Comparative proteomic analysis of resistant and susceptible cotton genotypes in response to leaf hopper infestation. Journal of Proteomics. 2024 Aug 15;305:105258.

[24] Gašo-Sokač D, Kovač S, Josić D. Application of proteomics in food technology and food biotechnology: process development, quality control and product safety.