European Journal of Biotechnology and Bioscience
ISSN: 2321-9122; Impact Factor: RJIF 5.44
Received: 10-09-2020; Accepted: 25-09-2020; Published: 12-10-2020
www.biosciencejournals.com
Volume 8; Issue 5; 2020; Page No. 33-36

Introduction to proteomics and its application in fisheries sector

Hamsavalli R[1], Suresh E[2], Deepak Agarwal[3], Kathirvelpandian A[4*]
[1-4] Institute of Fisheries Post Graduate Studies, TNJFU OMR Campus, Vaniyanchavadi, Chennai, Tamil Nadu, India

Abstract
"Proteome" refers to all proteins produced by a species at any particular point of time. The proteome varies with time and is defined as the proteins present in one sample (tissue, organism, cell culture) at a certain point of time. Various databases are used to study and analyze the proteome of the organism. Proteomics is a promising area of research with potential applications in the fisheries sector such as determination of fish health factors for various growth conditions, external influences, determining disease state and so on.

Keywords: proteomics, fisheries, protein databases

Introduction
Proteome analysis is a method to describe the molecular basis of (patho) physiological processes. Life is the translation of the static genome into highly dynamic proteomes. Proteome analysis supplements gene sequence data with protein information about where and in which ratio and under what conditions proteins are expressed. The terms 'proteome' or 'proteomics' was first introduced in the year 1995 [1]. The word 'proteome' was designed to denote the protein complement of a genome [2]. Simply, proteomics means the systematic analysis of protein profiles of tissues in an organism. Proteome refers to all the proteins produced by a species, much as the genome is the entire set of genes. Unlike the genome, the proteome varies with time and is defined as the proteins present in one sample (tissue, organism, cell culture) at a certain point of time. There are large numbers of proteomes present in an organism when compared with genome. For e.g. in humans there are 30,000 genes that can potentially encode 30,000 different proteins. However, alternative splicing & post translational modification may increase the number up to few million protein fragments or proteins. Proteomics at a broad level are the attempts to catalogue and characterize the proteins, compare variation in their expression levels in healthy and diseased condition, study their interaction and identify their functional role [3].

Protein information resource (PIR)
The protein information Resource (PIR) located at Georgetown University Medical Centre is an integrated public bioinformatics resource to support genomic and proteomic research and scientific study. PIR was established in 1984 by the National Biomedical Research Foundation as a resource to assist researchers in identification and interpretation of protein sequence information. PIR maintains the protein sequence Database (PSD), an annotated protein database containing over 2,83,000 sequences covering the entire taxonomic range. PIR has developed a bibliography system for literature searching, mapping and user submission etc.
PIR also maintains NREF, a Non- Redundant Reference database and iProclass, an integrated database for protein family, function, and structural information. PIR - NREF provides a timely and comprehensive collection of protein sequences with more than 10,00,000 entries from PIR-PSD, SWISS-PROT, RefSeq, GenPept and PDB etc. PIR website connects data analysis tools to underlying databases for information retrieval and knowledge discovery, with functionalities for interactive queries, combination of sequence and text search, sorting and visual exploration of search results. The FTP site provides free download for PSD and NREF biweekly release and auxiliary databases and files. Initially PIR was developed by Dr. Margaret Dayhoff as a collection of protein sequence for investigation of evolutionary relationship among proteins.

Primary protein sequence database
Since Primary structure of proteins mainly deals with the amino acid sequences, the primary database stores the amino acid sequences as linear alphabets that denotes the constituent amino acid residue. These usually have information about amino acid sequence in a peptide, which commonly have been derived as a result of original result. UniProt (Universal Protein Resource) is the world's most comprehensive and latest catalogue of information of proteins. It is central to repository of protein sequence and function, created by joining the information contained in Swiss-Prot, TrEMBL and PIR. PIR is being maintained collectively by NBRE, MIPS (Munich Centre for Protein Sequence) and JIPID (Japan International Protein Information Database) as PIR- International protein sequence database, since 1988. The data base endeavored to produce a High level annotation including description of the function of the proteins and the structure of the domains and then its post translational modifications etc., TrEMBL was created in 1996 as a computer supplement to Swiss-Prot and contains translation of all coding sequences in EMBL, a DNA sequence Database [4].

Primary structure analysis and prediction
Various tools are available for predicting the physical properties of protein by using the sequence information [5]. Some of the major tools are:
a. Compute pI/ Mw: A tool that calculates the isoelectric

point & molecular weight of an input sequence. The sequence input can be in the FASTA format, the output is the pI and molecular weight for the entire length of the protein sequence.

b. Peptide Mass: Cleaves one or more protein sequences from the SWISS - PROT and / or TrEMBL databases or a user - entered protein sequence with a chosen enzyme and computes the masses of the generated peptides. Also returns theoretical isoelectric point and mass values for the protein of interest.
c. Saps (Statistical Analysis of Protein Sequence): A tool to evaluate a wide variety of protein sequence properties by using statistical criteria. The output usually contains file name, sequence printout, compositional analysis, Charge distributional analysis, distribution of other amino acid types, Repetitive structures, multiplets, periodicity analysis, spacing analysis.
d. Prot Param: Tool which allows the computation of various physical and chemical parameters for a given protein stored in SWISS PROT or TrEMBL or user sequences.
e. Prot Scale: Used to calculate the hydrophobicity of protein. drawhca can be used to draw HCA (Hydrophobic cluster Analysis) plot of a protein sequence.
f. Pest and Pest Find: Proteins with intracellular half-lives of less than two hours are found to contain region rich in proline, glutamic acid, serine and threonine (PEST), these are called PEST region. PEST identifies possible PEST region in a submitted probe using the molecular fraction of the P, E, S, T components and the hydrophobicity index of the region. PEST find is a computer program used to determine whether a protein contains a PEST region.
g. Coils: Is a program that compares a sequence to a database of known parallel two stranded coiled-coils and derives a similarity score.
h. Paircoil: Predicts the location of coiled-coil region is amino acid sequences.
i. Multi coil Program: Predicts the location of coiled - coil region in amino acid sequences and classifies the prediction as dimeric or trimeric. This method is based on the pair coil algorithm.

Secondary protein databases

Secondary databases are called secondary because not only that they talk about the secondary structure but also they contain the results of analysis of the sequences in primary sources. They are also called as pattern databases. Very often the sequence of an unknown protein is too distantly related to any of the known sequence but it can be identified by the occurrence of a particular cluster a secondary structure elements (e.g. α-β-α or β-α-β) that is commonly called as a motif. An amino acid sequence in a protein is divided into conserved and variable region. Secondary protein databases are classified into three type mainly single motif, multiple motif, full domain alignment which are used for construction of the data bases [6].

Technologies used in proteomic study

1. 2D Gel electrophoresis: It is a method for the separation and identification of proteins in a sample by displacement in 2-Dimension.
2. ISOTOPE Coded Affinity Tags (ICAT): It is an alternative to SDS - PAGE using ICAT, relative protein abundance can be compared between two samples.
3. MASS–Spectroscopy: It can be used for protein identification. It separates proteins according to their mass to charge (m/z) ratios the process of ionization can be done by techniques like MALDI (Matrix assisted laser desorption / ionization) and ESI (electrospray ionization). Tandem Mass spectrometry (MS /MS) analyses and Peptide Mass Fingerprinting (PMF) are some other techniques

Specialized protein databases

1. Enzyme databases: Brenda database analysis enzyme properties, LIGAND database analysis enzyme reaction, EMP for studying enzymes metabolic pathway.
2. Two: Dimensional gel electrophoresis data databases - they are available at Expasy, like MultiIdent, Pep MAPPER, and PROWL etc.
3. Mass spectrometry protein databases: Ms-Fit, Ms-Tag, Ms-Seg, Ms-pattern, Ms-Bridge, Ms-Digest, Ms-Product, Ms-comp etc. Mascot is very popular tool used to analyze the MS data
4. Proteome databases: The proteome analysis databases has been set up to provide comprehensive statistical and comparative analyses of the predicted proteomes of fully sequenced organisms like YPD (yeast protein data base)
5. Merops: Merops database provides set of files called Famcards and clan cards, describing the individual families and clans
6. GPCR Db: This is a database of sequences and other data relevant to the biology of G- protein coupled receptors (GPCRs).
7. Databases for prediction of post-translational Modification (PSORT): Used for prediction of protein sorting signal and localization sites signal IP for prediction of signal peptide cleavage sites others includes chlorop, Targets, Mitoprot II, Predotar, NetOGlyc, Netphos, Find Mod etc.

EST sequences databases

An expressed sequence tag (EST) is a short sequence of DNA taken from a DNA copy of mRNA. The EST is complementary to the m-RNA and can be used to identify genes corresponding to mRNA, ESTs are very useful for the construction of physical map of genome. Mendel - ESTs is EST data base created to apply controlled annotation to plants ESTs. Mendel - ESTs is primarily a database of plant ESTs, which have been compared to Mendel - GFDb, completely sequenced genomes and domain databases. This approach associates ESTs with individual sequences and the controlled annotation of gene families and protein domains [7].

The COGEME phytopathogenic fungi and Oomycetes EST database contains expressed sequence tags (ESTs) from eleven species of fungi & two Oomycetes specie [8]. In order to reduce sequence redundancy, ESTs representing the same gene were joined together in to a single contig or unisequence. Unisequences have been annotated based on similarity to known sequences and assigned to a functional classification group. It uses BLAST Algorithms.

Application of proteomics in fisheries sector

a. Proteome analysis can be highly focused in the investigation of specific protein complexes or concerned with profiling and comparisons of multiple constituents, with a view to relating the composition of the proteome to factors such as growth conditions, external influences, disease state, and so on. It can assist in the confirmation of DNA open reading frames, in the characterization of knockout mutants and in the validation of drug and vaccine targets discovery [1].
b. Proteome maps are also useful reference points for future studies on temporal or spatial protein synthesis in identifying sex- or stage-specific proteins, excretory/secretory proteins and quantifying stress responses [9]. Proteomics offers a major new approach to drug and vaccine discovery.
c. Proteomics is used for the differential characterization of commercially important species of fish at the molecular level, e.g. the proteomics technology has been applied for the first time to achieve differential characterization of the sarcoplasmic protein fraction of fish species. The identification of species-specific proteins was carried out by combining proteomic tools such as Matrix Assisted Laser-Desorption/ Ionization Time of Flight-Mass Spectrometry (MALDI-TOF-MS) and nasospray-ion trap-MS (nESI-IT-MS) with electrophoretic techniques [10].
d. Proteomics technology would help in fisheries for identification of serum/plasma proteins that might be involved in the constitutive resistance to infections, characterization of fish sarcoplasmic polypeptides/ muscle protein for species identification, biochemical analysis of cross-reactive antigens, understanding the molecular pathogenesis and the genetics of disease resistance [11].
e. Proteomics studies will be helpful for identification of proteins responsible for commercially important traits such as natural resistance to infections, faster growth rate, etc. This might help in identification of proteins and development of transgenic fishes with such important traits. Such studies may provide clues and aid in the development of immunodiagnostics and vaccines in the long run [1].
f. Fish Proteomics and Fish Protein Databases (FPD) might help in better understanding of phylogenetic position, ecological habitat, molecular pathogenesis, development of markers for marker-assisted selection and breeding, species/strain identification and raciation studies that will be useful for fish biodiversity assessment and conservation [1].
g. Species Authentication - Specific protein biomarkers are detectable with MALDI-TOF MS which could discriminate among several commercially important fish species such as those belonging to Gadidae and Pleuronectiformes [12].
h. Identification and Characterization of Allergens - Allergic reactions to seafood affect a significant part of the population: about 0.5% of young adults are allergic to shrimp [13]. Seafood allergies are caused by an immunoglobulin E-mediated response to particular proteins, including structural proteins such as tropomyosin [14]. Proteomics provide a highly versatile toolkit to identify and characterize allergens [15].
i. Proteomic approaches in fish meat quality research – Few researchers have undertaken proteome research related to fish/seafood color attributes, although various pre-harvest and post-harvest factors could influence muscle food quality of fish and seafood [16].

Conclusion

Fisheries and aquaculture are important food production sectors which contribute to food and nutritional security. In the last decade, proteomic technologies have been increasingly used in fish biology research [17]. Proteomics has been applied primarily to investigate the physiology, nutrition, health, quality and food safety, development biology and the impact of contaminants in fish model organisms, such as zebrafish (Danio rerio), as well as in some commercial species produced in aquaculture, mainly salmonids and cyprinids. However, the lack of previous genetic information on most fish species has been a major drawback for a more general application of the different proteomic technologies which currently available [18].

References

1. Wilkins MR, Sanchez JC, Gooley AA, Appel RD, Humphrey-Smith I, Hochstrasser DF, *et al.* Biotechnol. Gene. Eng. Rev. 1995; 13:19.
2. Mohanty S, Mohanty S, Panwar RS. Proteomics & its Potential Application in Fisheries Sector. Fishing Chimes, 2005, 2(6).
3. Wilkins MR, Williams KL, Apple RD, Hochstrasser DF. Proteome Research: New Frontiers in Functional Genomics. Springer, Berlin, 1997; 1-243.
4. Barker WC, Garavelli JS, Huang H, Mc Garvey PB, Orcutt BC, Srinivasarao GY, *et al.* The protein information resource (PIR). Nucleic acids research. 2000; 28(1):41-44.
5. Eidhammer I, Jonassen I, Taylor WR. Protein Bioinformatics: An algorithmic approach to sequence and structure analysis. Chichester: Wiley. 2004, 1.
6. Rost B. Protein secondary structure prediction continues to rise. Journal of structural biology. 2001; 134(2-3):204-218.
7. Picoult-Newberg L, Ideker TE, Pohl MG, Taylor SL, Donaldson MA, Nickerson DA, *et al.* Mining SNPs from EST databases. Genome Research, 1999; 9(2):167-174.
8. Xu J, Linning R, Fellers J, Dickinson M, Zhu W, Antonov I, *et al.* Gene discovery in EST sequences from the wheat leaf rust fungus Puccinia triticina sexual spores, asexual spores and haustoria, compared to other rust and corn smut fungi. BMC genomics. 2011; 12(1):161.
9. Barrett JR, Jefferies JR, Brophy PM. Parasitol. Today. 2000; 16:4000.
10. Pineiro C, Vazquez J, Marina AI, Barros-Velazquez J, Gallado JM. Electrophoresis. 2001; 22:1545.
11. Verrez-Bagnis V, Ladrat C, Morzel M, Noel J, Fleurence J. Electrophoresis. 2001; 22:1539.
12. Mazzeo MF, Giulio BD, Guerriero G, Ciarcia G, Malorni A, Russo GL, *et al.* Fish authentication by MALDI-TOF mass spectrometry. Journal of agricultural and food chemistry. 2008; 56(23):11071-6.
13. Woods RK, Thien F, Raven J, Walters EH, Abramson M. Prevalence of food allergies in young adults and their relationship to asthma, nasal allergies, and eczema. Annals of Allergy, Asthma and

Immunology. 2002; 88(2):183-189.
14. Lehrer SB, Ayuso R, Reese G. Seafood allergy and allergens: a review. Marine Biotechnology. 2003; 5(4):339-348.
15. Sveinsdóttir H, Martin SA, Vilhelmsson OT. Application of proteomics to fish processing and quality. Food Biochemistry and Food Processing, 2012, 406.
16. Joseph P, Nair MN, Suman SP. Application of proteomics to characterize and improve color and oxidative stability of muscle foods. Food Research International, 2015; 76:938-945.
17. Rodrigues PM, Silva TS, Dias J, Jessen F. Proteomics in aquaculture: applications and trends. Journal of proteomics. 2012; 75(14):4325-4345.
18. Mohanty BP, Samanta S, Hassan MA, Behera BK, Pandit A, Manna RK, *et al.* Book of Abstracts: National Seminar on "Priorities in Fisheries and Aquaculture"(PFA-2017). ICAR–Central Inland Fisheries Research Institute, Barrackpore, Kolkata-700120, India, 2017, 01-227.