Journal of
Computer Science & Systems Biology

Nanda et al. J Comput Sci Syst Biol 2011, S13
DOI: 10.4172/jcsb.S13-002

Review Article Open Access

Integration of Bioinformatics Tools for Proteomics Research

Tushar Nanda[1]*, Kadambini Tripathy[1] and Ashwin P[2]

[1]Department of Bioscience and Biotechnology, Fakir Mohan University, Odisha-756020, India
[2]Department of Biotechnology, Vinayaka Missions University, Tamil Nadu, India

Abstract

An emerging field for the analysis of biological systems is the study of the complete protein complement of the genome, the 'proteome'. Proteomics is defined as a scientific approach used to elucidate all protein species within a cell or tissue. Emerging proteomic technologies have promised for early diagnosis and in advancing treatment directions. Application of these technologies has produced new biomarkers, diagnostic approaches, and understanding of disease biology. Here in this review we have outlined potential implications for clinical proteomics focused on applied research activities. This review presents a general survey of the recent development in array technologies from a proteomics perspective. The significance of informatics in proteomics will gradually increase because of the advent of high-throughput methods relying on powerful data analysis. Lastly, an attempt has been made to present novel biological entities named the bioinformatics tools developed to analyse the large protein–protein interaction networks they form, along with several new perspectives of the field. Bioinformatics is an integral part of proteomics research.

Keywords: Proteome; Biomarker; Protein; Chromatography

Introduction

Proteomics and Bioinformatics are rooted in life sciences as well as computer and information sciences and technologies. Both of these interdisciplinary approaches draw from specific disciplines such as mathematics, physics, computer science and engineering, biology and behavioral science. Proteomics and bioinformatics each maintain close interaction with life sciences to realize their full potential. Now a days proteomics without bioinformatics is just like a boat without radar. Bioinformatics uses computational approaches to address theoretical as well as experimental questions in biology. The growth of the biotechnology industry in recent years is unprecedented and advancements in molecular modeling, disease characterization, pharmaceutical discovery, clinical healthcare, forensics, and agriculture fundamentally impact economic and social issues worldwide [1].

Scientific research has changed in recent years mainly due to the completion of numerous genomes in addition to the development and application of high-throughput technologies including gene expression microarrays and mass spectrometry. The development of all computational tools have not only gave a ray of hope but also has imposed new challenges because of the large amount of data that needs a comprehensive analysis, but at the same time it has also brought about new and increasing opportunities to enhance the knowledge on biological systems as a whole [2].

Proteins are the main catalysts, structural elements, signaling messengers and molecular machines of a cell, which is a strong argument to support the advantages and importance of directly analyzing proteins. Proteomics is defined as the large scale identification and functional characterization of all expressed proteins in a given cell (in a given state), including all protein isoforms and modifications, protein interaction networks, protein structure determination and high order complexes of proteins. An important progress in proteomics has been achieved by the introduction of powerful new technologies and high throughput experiments and the integration of bioinformatics tools to analyze the results of those experiments. Several reviews address the advancing technology available for proteomic studies [2]. Experimental methodologies, fine-tuned in recent years to allow high-throughput protein and cDNA analyses, have resulted in exponential growth of protein and cDNA expression profiles and interaction datasets. A number of large-scale analyses, such as the two-hybrid interaction maps and cDNA microarray technology, now allow interaction and expression datasets from large numbers of genes to be analyzed quickly and efficiently in a single experiment. Protein profiling arrays for the comparable large-scale analysis of protein expression patterns are under active development as well. When perfected, their output should be equally prolific. Finally, mass spectrometry, possibly the most important proteomics tool to date, generates vast quantities of data through large-scale liquid chromatography (LC), tandem mass spectrometry (MS/MS) for identification of expressed proteins in complex mixtures [3].

Integrative Approach: An Overview

Proteomics is the large-scale study of proteins, particularly their structures and functions. Proteins are vital parts of living organisms, as they are the main components of the physiological metabolic pathways of cells. Proteomics has shown a great achievement since a decade or long. It has rather a more complicated system than genomics. The complexity arrived when proteome differed from cell to cell and from time to time.

Since a decade or two MS has been proved to be an outstanding technology for the analysis of complex mixtures along with 2-D gel electrophoresis. This technology is being applied in fields like clinical sciences, medical sciences, environmental, life sciences, engineering etc. Various approaches are being discovered day by day which are highlighted. A proteomic-based approach was applied to characterize cellular responses of neuronal cells to Pyridostigmine Bromide

***Corresponding author:** Tushar Nanda, Department of Bioscience and Biotechnology, FM University, Odisha, E-mail: ertusharnanda@gmail.com

Received April 15, 2011; **Accepted** May 20, 2011; **Published** May 23, 2011

Citation: Nanda T, Tripathy K, Ashwin P (2011) Integration of Bioinformatics Tools for Proteomics Research. J Comput Sci Syst Biol S13. doi:10.4172/jcsb.S13-002

exposure. Protein extracts from cultured neuroblastoma cells treated with 700 nM PB for 10 days, as well as extracts from control cells were separated using two-dimensional gel electrophoresis. Twenty two differentially-expressed proteins were identified by MALDI-TOF mass spectrometry (MS) [5]. Similarly Maldi-TOF MS was applied to identify the affected proteins when exposed to 1800 MHz GSM mobile phone [6]. The use of Conjunctival Swab for the proteomic characterization of dry eye syndrome utilizes clinically based non invasive methodology for collection of specimen from the posterior lid and inferior conjunctival mucosa of the subject [9]. The 2-D gel electrophoresis is useful for the analysis of plasma proteins leading to biomarker discovery of human diseases. Results showed that the depletion of two high-abundant proteins improved the visualization of less abundant proteins present in human plasma and precipitation with TCA/acetone resulted in an efficient sample concentration and desalting. We also found that visualization of 2D gel profiles by silver staining and fluorescent staining enhanced the detection of low abundant plasma proteins as compared to Coomassie staining [7].

In an approach the automated calculation of unique peptide sequences has been done in two steps: In a first step a SQL-based database of theoretically digested peptides from a given FASTA file formatted protein database is generated by choosing a protease. In a second step, *in silico* generated peptides from a pre-defined protein sequence are compared to this peptide database in order to identify unique peptides [8]. Proteome Analysis of Serum-Containing Conditioned Medium from Primary Astrocyte Cultures [10] showed that this is the first study to identify secreted proteins in serum containing medium using a proteomic approach involving stable isotope labeling by amino acids in cell cultures and mass spectrometry. Serum proteome analysis provides a potential promising approach in disease diagnosis and therapeutic monitoring [11]. The removal of high-abundant proteins in serum by the ProteoExtract™ Albumin Removal column, ethanol precipitation, the heating with 2.5% SDS and 2.3% DTT to denature sample at 95°C for 3 min, and IEF on pH 4-7 IPG strips (17cm) with 100 μg depleted serum proteins are generally recommended for serum proteome analysis on 2-DE by silver staining, which can effectively improve the resolution and intensity of low-abundant proteins.

Pathway modeling is one of the most interesting as well as new aspects of systems biology to design and analyze pathway for various diagnostic as well as other purposes [12]. Integration and prediction of protein protein interactions with the help of a PPI software framework "PIANA" solves many of the nomenclature issues common to systems dealing with biological data [13]. One of the recent visualization tool e.g. DataBiNS-Viz enables execution of the DataBiNS workflow on proteins described by KEGG, PubMed, or OMIM identifiers, followed by manual exploration of the integrated structure/function and pathway data for those proteins, with a particular focus on nsSNP data in-context [14]. The homology modeling by using MODELLER 9v2 software was done to predict a 3-Dimensional structure of Cathepsin L Protein which degrades connective tissue proteins like collagen, elastin and fibronectin. The final model obtained by molecular mechanics and dynamics method and was assessed by PROCHECK and VERIFY 3D graph, which showed that the final refined model is reliable [15]. Similarly Global Proteomics is an alternate approach where all blood proteins modulated by disease or drug are used to resolve pharmacodynamic questions without the time, cost, and risk of developing an immunoassay [16]. The use of PupaSuite, UTRscan and miRBase computational tools as a pipeline for the prediction of miRNA and their target evaluated the functional role of mRNA in colon cancer [17]. An integrative visualization methodology combining experimentally produced proteomic features with protein meta-features is found to be effective in filtering by using a software called VIP software for it [24]. The application of magnetic bead-based purification (ClinProt system) followed by matrix-assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF-MS) to profile human tear proteins is ideally suited for the first line screening of peptides and proteins in tears [18]. Electron transfer dissociation (ETD) of peptide ions has been observed as a better tool for mass spectrometry based peptide sequencing than collision Induced Dissociation (CID) [19]. A 2-Dimensional electrophoresis step followed by western blotting detection and MS identification was done to profile phosphorylated proteins in Human fetal liver (HFL) aged 16-24 weak of gestation [21] where low degree serine, threonine and tyrosine were found when proteins associated with hematopoiesis. Imaging mass spectrometry (IMS) is an emerging technology which uses matrix deposition and MALDI-TOF mass spectrometry instrumentation for image generation [22] with the application for murine brain tissues.

The Informatics Prone

Bioinformatics is appearing as an empowering technology in different fields. Analysis of protein-protein interaction, protein identification and characterization, post translational modification prediction, primary- secondary- tertiary and quaternary structure prediction, molecular modeling and visualization, Prediction of disordered regions and other analysis with this emerging technology-is the hub of present bioinformatics in proteomics as well as glycomics and metabolomics research.

While several techniques are available in proteomics, LC -MS based analysis of complex protein/ peptide mixtures have turned out to be a main stream analytical technique for quantitative proteomics [4].

Due to the large variety of Proteomics workflows, as well as the large variety of instruments and data-analysis software available Human Proteome Organisation (HUPO) [23] is facing a major challenge and expects lead to field-generated data of greater accuracy, reproducibility and comparability where a new generation of the ProteinScape™ bioinformatics platform, now enabling researchers to manage Proteomics data from the generation and data warehousing to a central data repository with a strong focus on the improved accuracy, reproducibility and comparability of protein datas.

The homology model of Late Embryogenesis Abundant protein was generated by using the LOOPP software. The final model obtained by molecular mechanics and dynamics method was assessed by PROCHECK. The model could prove useful in further functional characterization of this protein [20]. Coon OMSSA Proteomic Analysis Software Suite (COMPASS) is a free and open-source software pipeline for high-throughput analysis of proteomics data, designed around the Open Mass Spectrometry Search Algorithm [29]. The SPIRE (Systematic Protein Investigative Research Environment) provides web-based experiment-specific mass spectrometry (MS) proteomics analysis. Its emphasis is on usability and integration of the best analytic tools which provides an easy to use web-interface and generates results in both interactive and simple data formats [42]. The superior performance of ScanRanker enables it not only to find unassigned high quality spectra that evade identification through database search but also to select spectra for de novo sequencing and cross-linking analysis of proteins [41].

Novel and improved computational tools are required to transform large-scale proteomics data into valuable information of biological relevance [45]. Proteo Connections, a bioinformatics platform tailored to address the pressing needs of proteomics analyses. Similarly the Pathway Browser is a Systems Biology Graphical Notation (SBGN)-like visualization system that supports manual navigation of pathways by zooming, scrolling and event highlighting, and that exploits PSI Common Query Interface (PSIQUIC) web services to overlay pathways with molecular interaction data from the Reactome Functional Interaction (FI) Network and interaction databases such as IntAct, ChEMBL, and BioGRID [47]. The ProteoRed MIAPE web toolkit, a new web-based software suite that performs several complementary roles related to proteomic data standards [51]. Proteomic repositories like PRIDE, The Global Proteome Machine, PeptideAtlas etc. are available to harbor the wealth of mass spectral data and appropriate protein identifications [52]. ANTILOPE, a new fast and flexible approach based on mathematical programming combines Lagrangian relaxation for solving an integer linear programming formulation with an adaptation of Yen’s k shortest paths algorithm [50]. Using a classical *in silico* technique, we have identified four best epitopes from three transporters are found to be antigenic and predicted to induce both the T- and B-cell mediated immune [56]. Peptidomimetics Based Inhibitor Design is drug designing strategy mimicking the framework of a short peptide [57]. Genome Medicine Database of Japan Proteomics (GeMDBJ Proteomics is a freely accessible database [59]. The HUPO proteomics database utilises with XML formats to exchange and import data into databases, allowing direct access and comparability irrespective of the originating instrumentation [60]. Primer designing for cold induced gene, DREB1A is done using Primer3 software [61]. 74 proteins that are likely to be involved in Alzheimer's disease by employing multiple sequence alignment using ClustalW tool and constructed a Phylogenetic tree using functional protein sequences extracted from NCBI [62]. Another approach to compare different 2-D gels is Web server AC2Dgel [63]. Comparative Modeling Of Viral Protein R (Vpr) From Human Immunodeficiency Virus 1 (Hiv 1) is done to predict the structure of a VpR using NMR structure of the HIV-1 regulatory protein as the template (PDB ID: 1ESX: A) [66]. Proteolytic enzyme database includes enzyme's class, source, EC, molecular weight, N-terminal, C-terminal ,thiols, activators, inhibitors, bond specificity and comments. This database will be of high values for researchers and students working in this area [67]. *In Silico* Prediction of Epitopes in Virulence Proteins of *Mycobacterium Tuberculosis* H37Rv for Diagnostic and Subunit Vaccine Design [68] is done. siRNA Scanner uses a fuzzy logic-based system to calculate siRNA qualities. This program is fully built in Practical Extract Report Language (PERL 5.8.8.6 Build 820) and accessible in a command line interface. siRNA Scanner's high performance, minimal user interaction, and its fast algorithm, make this program useful for selecting Small Interfering RNA for gene expression studies [69]. π-calculus and the following up features of Stochastic Pi Machine (SPiM) programming language allows to change the topology of genetic networks which resembles a gene exchange in nature [74]. Here the author used five elements such as decay, null gate, gene product, negative and positive gates.

Recent Developments in Proteomics

Liquid chromatography-mass spectrometry (LC/MS) and gel electrophoresis-mass spectrometry (2-DE/MS) have been applied to investigate the proteome of a number of lung-origin samples, including sputum, bronchoalveolar lavage fluid, exhaled-breath condensate to discover and validate biomarkers as prognostic tools of development and progression of the Chronic obstructive pulmonary disease (COPD) [25]. Quantitative proteomics based on 2D electrophoresis (2-DE) coupled with peptide mass fingerprinting is still one of the most widely used quantitative proteomics approaches in microbiology research. The particular focus is given to the emerging field of toxicoproteomics, a new systems toxicity approach that offers a powerful tool to directly monitor the earliest stages of the toxicological response by identifying critical proteins and pathways that are affected by, and respond to, a chemical stress [26]. The identification, quantitation and global characterization of all proteins within a given proteome are extremely challenging due to the absolute detection limits of technology as well as the dynamic range in expression of proteins; and the extreme diversity and heterogeneity of the proteome. A range of technologies highlighting the challenges of protein and peptide analysis in the context of proteome research and some of the advantages and disadvantages of present techniques are reviewed [27]. Activity Based Protein Profiling, Trifunctional Capture Compounds and affinity chromatography are the basis for a generic approach are the three approaches to reduce the proteome complexity on the basis of functional small molecule-protein interactions [28]. Multidimensional LC-MS data sets have been demonstrated to identify and quantitate 2000-8000 proteins from whole cell extracts. The analysis of the resulting data sets requires several steps from raw data processing, to database-dependent search, statistical evaluation of the search result, quantitative algorithms and statistical analysis of quantitative data [30].

The proteomic approaches, namely two-dimensional electrophoresis (2DE), mass spectrometry (MS), and bioinformatics tools used in our recent studies to gain insight into the proteins potentially involved in low-temperature or mutagenic treatment-induced rescue process of F508del-CFTR [31]. The various statistical methods and models for bridging omics data levels have been described [33] from the fields of Machine Learning and Pattern Recognition with particular focus on transcriptomics and proteomics profiles. Protein microarrays are a high-throughput technology capable of generating large quantities of proteomics data by using essential algorithms, including spot-finding on slide images, Z score, and significance analysis of microarrays (SAM) calculations, as well as the concentration dependent analysis (CDA) [34]. Advances in mass spectrometry-driven proteomics rely on robust bioinformatics tools that enable large-scale data analysis [36]. Using high-throughput body fluid profiling by MALDI-TOF mass spectrometry, small proteins and peptides were detected as promising candidate biomarkers for diagnosis and disease progression of MS [38]. With the aid of in silico predictive tools, a unique 9-mer tryptic peptide of P-glycoprotein protein was synthesized (with the stable isotope labeled (SIL) peptide as internal standard) and applied for quantitative LC/MS/MS method development [39]. The treatment of biological samples with solidphase combinatorial peptide ligand libraries enables detection of rare species [58]. A new resource, RedundancyMiner, that de-replicates the redundant and nearly-redundant GO categories that had been determined by first running GoMiner. The main algorithm of RedundancyMiner, MultiClust, performs a novel form of cluster analysis in which a GO category might belong to several category clusters. Each category cluster follows a "complete linkage" paradigm. The metric is a similarity measure that captures the overlap in gene mapping between pairs of categories [32].

Most common methods for the systematic identification of protein interactions and exemplify different strategies for the generation of protein interaction networks focusses on the recent development of protein interaction networks derived from quantitative proteomics

data sets [40]. High-throughput genomic sequencing and quantitative mass spectrometry (MS)-based proteomics technology have recently emerged as powerful tools, increasing our understanding of chromatin structure and function [43]. Lectin microarray is an emerging technique enabling multiplex glycan profiling in a direct, rapid and sensitive manner [65]. The Proteogenomic Mapping Tool provides a standalone application for mapping peptides back to their source genome on a number of operating system platforms with standard desktop computer hardware and executes very rapidly for a variety of datasets [37]. The application of novel methods for identifying S-nitrosylated proteins, especially when combined with mass-spectrometry based proteomics to provide site-specific identification of the modified cysteine residues, promises to deliver critical clues for the regulatory role of this dynamic posttranslational modification in cellular processes [44]. Visualization methods for detection of proteins separated by polyacrylamide gel electrophoresis. uses staining approaches involve colorimetric dyes such as Coomassie Brilliant Blue, fluorescent dyes including Sypro Ruby, newly developed reactive fluorophores, as well as a plethora of others [46]. A microsensor technique based on phase sensitive detection of real time biophysical transport is reviewed [48]. classification of the identified proteins into their functional categories indicated that Side Population cells over express stress proteins, cytoskeletal proteins and enzymes of the glycolytic metabolism [53]. The degree of automation in proteomics has yet to reach that of genomic techniques, but even current technologies make a manual inspection of the data infeasible. The key algorithmic problems bioinformaticians face when handling modern proteomic samples and shows common solutions to them [35]. A novel mass spectrometry (MS)-based approach to the identification of host-derived biomarkers (BMs) in the circulating low-molecular-mass serum proteome was tested [54]. RRM2 is a downstream target of the ATM-p53 pathway that mediates radiation-induced DNA repair. We demonstrated that the integrated bioinformatics approach facilitated pathway analysis, hypothesis generation and target gene/protein identification [64]. Double exponential model [64] is a recent technique discovered to analyse protein interaction network. A mass spectrometry-based **proteomics** strategy to examine protein-protein interactions using anti-Green Fluoroscent Protein single-chain antibody V(H)H in a combination with a novel stable isotopic labeling reagent, isotope tag on amino groups (iTAG) [49]. The effects of shear stress on the mechanosensing and protein phosphorylation pathways were altered by HG which were validated by Western blot analysis, suggesting that HG importantly modulates shear stress-mediated endothelial function [55]. Damage analysis is of higher priority these days which utilizes various statistical methods e.g. graph analysis to study the damages in various metabolic pathways [75].

Bioinformatics as an R & D Tool

Rapid advances in technologies like genomic as well as bioinformatics coupled with a unique collaboration between industry and academia are beginning to show the true potential for the human genome project to affect patient healthcare. By knowing the sequence of the human genome and beginning to unravel the location and sequence of all genes and their variants, scientists can establish a better understanding of the mechanisms for diseases, with subsequent availability of new treatments. Because of the vast amount of data coming out of the Human Genome Project, bioinformatics tools and databases have become an integral part of pharmacogenomic and disease susceptibility gene research. They play an important role in candidate gene identification, gene finding, SNP detection, genotyping and genetic analysis. Bioinformatics data integration and tool standardization are critical to the success of association and linkage studies. The underlying data models accommodate the variability inherent in subject collections, the ability to trace the data source, and the automation and archival storage of analysis results. A fully traceable data source is important, as we are often faced with anomalies in data at a late stage that can be very time consuming to resolve in an infrastructure that does not facilitate data integration. The polymorphism database component includes data from public and proprietary sources [70].

Tools Integrated in R & D Sector [72]

The emerging tools used in various R& D sectors are summarized as below Protein identification and characterization with peptide mass fingerprinting data

FindMod - Predict potential protein post-translational modifications and potential single amino acid substitutions in peptides. Experimentally measured peptide masses are compared with the theoretical peptides calculated from a specified Swiss-Prot entry or from a user-entered sequence, and mass differences are used to better characterize the protein of interest.

FindPept - Identify peptides that result from unspecific cleavage of proteins from their experimental masses, taking into account artefactual chemical modifications, post-translational modifications (PTM) and protease autolytic cleavage

Mascot - Peptide mass fingerprint from Matrix Science Ltd., London

PepMAPPER - Peptide mass fingerprinting tool from UMIST, UK

ProFound - Search known protein sequences with peptide mass information from Rockefeller and NY Universities [or from Genomic Solutions]

ProteinProspector - UCSF tools for peptide masses data (MS-Fit, MS-Pattern, MS-Digest, etc.)

Identification with isoelectric point, molecular weight and/or amino acid composition

AACompIdent - Identify a protein by its amino acid composition

AACompSim - Compare the amino acid composition of a UniProtKB/Swiss-Prot entry with all other entries

TagIdent - Identify proteins with isoelectric point (*pI*), molecular weight (*Mw*) and sequence tag, or generate a list of proteins close to a given *pI* and *Mw*

MultiIdent - Identify proteins with isoelectric point (*pI*), molecular weight (*Mw*), amino acid composition, sequence tag and peptide mass fingerprinting data

Pattern and profile searches

InterPro Scan - Integrated search in PROSITE, Pfam, PRINTS and other family and domain databases

MyHits - Relationships between protein sequences and motifs

ScanProsite - Scans a sequence against PROSITE or a pattern against the UniProt Knowledgebase (Swiss-Prot and TrEMBL)

HamapScan - Scans a sequence against the HAMAP families

MotifScan - Scans a sequence against protein profile databases (including PROSITE)

Pfam HMM search - Scans a sequence against the Pfam protein families db [At Washington University or at Sanger Centre]

ProDom - Compares sequences with ProDom search utility

SUPERFAMILY Sequence Search - Assign SCOP domains to your sequences using the SUPERFAMILY hidden Markov models

FingerPRINTScan - Scans a protein sequence against the PRINTS Protein Fingerprint Database

3of5 - Complex Pattern Search - e.g. to search for a motif with 3 basic AA in 5 positions

ELM - Eukaryotic Linear Motif resource for functional sites in proteins

PRATT - Interactively generates conserved patterns from a series of unaligned proteins; [at EBI / ExPASy]

Post-translational modification prediction

ChloroP - Prediction of chloroplast transit peptides

LipoP - Prediction of lipoproteins and signal peptides in Gram negative bacteria

MITOPROT - Prediction of mitochondrial targeting sequences

PATS - Prediction of apicoplast targeted sequences

PlasMit - Prediction of mitochondrial transit peptides in Plasmodium falciparum

Predotar - Prediction of mitochondrial and plastid targeting sequences

PTS1 - Prediction of peroxisomal targeting signal 1 containing proteins

SignalP - Prediction of signal peptide cleavage sites

DictyOGlyc - Prediction of GlcNAc O-glycosylation sites in Dictyostelium

NetCGlyc - C-mannosylation sites in mammalian proteins

Primary structure analysis

ProtParam - Physico-chemical parameters of a protein sequence (amino-acid and atomic compositions, isoelectric point, extinction coefficient, etc.)

Compute pI/Mw - Compute the theoretical isoelectric point (*pI*) and molecular weight (*Mw*) from a UniProt Knowledgebase entry or for a user sequence

ScanSite pI/Mw - Compute the theoretical *pI* and *Mw*, and multiple phosphorylation states

MW, pI, Titration curve - Computes *pI*, composition and allows to see a titration curve

Scratch Protein Predictor

HeliQuest - A web server to screen sequences with specific alpha-helical properties

Radar - De novo repeat detection in protein sequences

REP - Searches a protein sequence for repeats

REPRO - De novo repeat detection in protein sequences

Tertiary structure prediction

Homology modelingSWISS-MODEL - An automated knowledge-based protein modelling server

CPHmodels - Automated neural-network based protein modelling server

ESyPred3D - Automated homology modeling program using neural networks

Geno3d - Automatic modelling of protein three-dimensional structure

Threading

Phyre (Successor of 3D-PSSM) - Automated 3D model building using profile-profile matching and secondary structure

Fugue - Sequence-structure homology recognition

HHpred - Protein homology detection and structure prediction by HMM-HMM comparison

LOOPP - Sequence to sequence, sequence to structure, and structure to structure alignment

SAM-T08 - HMM-based Protein Structure Prediction

PSIpred - Various protein structure prediction methods (including threading) at Bloomsbury Centre for Bioinformatics

Quaternary structure

MakeMultimer - Reconstruction of multimeric molecules present in crystals

EBI PISA - Protein Interfaces, Surfaces and Assemblies

PQS - Protein Quaternary Structure Query form at the EBI

ProtBud - Comparison of asymmetric units and biological units from PDB and PQS

Molecular modeling and visualization tools

Swiss-PdbViewer - A program to display, analyse and superimpose protein 3D structures

SwissDock - Docking of small ligands into protein active sites with EADock DSS

SwissParam - Topology and parameters for small molecules, for use with CHARMM and Figure 1.

Updates on Current Projects Going on [71]

- ✓ Identification of proteins in placental blood associated with spontaneous labour.
- ✓ Proteomic Analysis of the nucleolus following viral infection.
- ✓ Use of proteomics to identify plasma factors that affect podocyte differentiation
- ✓ Identification of Immuno-reactive species
- ✓ Identification of substrates of Protein Kinase B
- ✓ Identification of potential allergens in midge salivary glands.

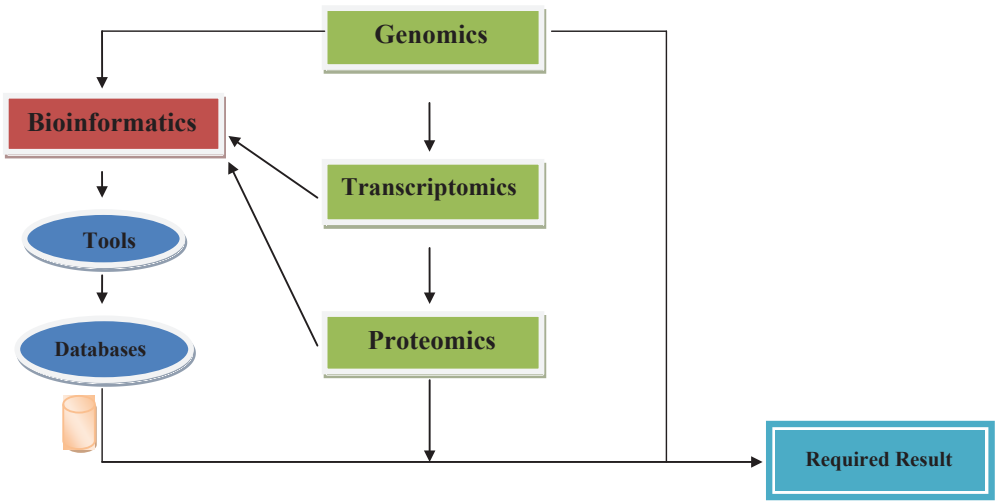

Figure 1: An integrative approach of bioinformatics and proteomics .

- ✓ Characterisation of two antibodies to oligomeric Aß and their use in ELISAs on human brain tissue homogenates.
- ✓ Analysis of mitochondrial protein phosphorylation

Aims of Proteomics [73]

- Detect the different proteins expressed by tissue, cell culture, or organism using 2-Dimensional Gel Electrophoresis
- Store those information in a database
- Compare expression profiles between a healthy cell Vs. a diseased cell
- The data comparison can then be used for testing and rational drug design.

Future Prospects

Although proteomics has proved its promise for biomarker discovery, further work is still required to enhance the performance and reproducibility of established proteomics tools before they can be routinely used in the clinical laboratory. Issues regarding pre-analytical variables, analytical variability and biological variation must be tackled. The efforts of many researchers, together with the HUPO consortium, are now starting to address many of these issues, such that the future remains extremely bright for the widespread use of proteomics in different fields. Further the integration of proteomics and bioinformatics will not only yield a satisfactory result but also help to gather information for future purposes.

Conclusion

Recent progress in biotechnology offers the promise of better medical care at lower costs. Among the techniques that show the greatest promise is mass spectrometry of proteins, which can identify proteins present in body fluids and tissue specimens at a large scale. The field of proteomics is expanding rapidly due to the completion of the human genome and the realization that genomic information is often insufficient to comprehend cellular mechanisms. Interest in mass spectrometry and its use as a new clinical diagnostic tool has grown rapidly and is poised to become an important medical field for the next century. Proteomics methods are essential for studying protein expression, activity, regulation and modifications. The recent developments and applications in proteomics are discussed including mass spectrometry data analysis and interpretation, analysis and storage of the gel images to databases, gel comparison, and advanced methods to study e.g. protein co-expression, protein–protein interactions, as well as metabolic and cellular pathways.

References

1. Rao VS, Das SK, Rao VJ and Gedela S (2008) Recent developments in life sciences research: Role of Bioinformatics. African Journal of Biotechnology 7: 495-503.
2. Ulloa AR, Rodríguez R (2008) Bioinformatic tools for proteomic data analysis: an overview. Biotecnología Aplicada 25: 312-319.
3. Lundgren DH, Eng§ J, Wright§ ME, and Han DK (2003) PROTEOME-3D: An Interactive Bioinformatics Tool for Large-Scale Data Exploration and Knowledge Discovery. Molecular & Cellular Proteomics 2:1164–1176.
4. Tuli L, Ressom HW (2009) LC–MS Based Detection of Different ial Protein Expression. J Proteomics Bioinform 2: 416-438.
5. Abdullah L, Reed J, Kayihan G, Mathura V, Mouzon B, et al. (2009) Proteomic Analysis of Human Neuronal Cells Treated with the Gulf War Agent Pyridostigmine Bromide. J Proteomics Bioinform 2: 439-444.
6. Nylund R, Tammio H, Kuster N, Leszczynski D (2009) Proteomic Analysis of the Response of Human Endothelial Cell Line EA.hy926 to 1800 GSM Mobile Phone Radiation. J Proteomics Bioinform 2: 455-462.
7. Ahmad Y, Sharma N (2009) An Effective Method for the Analysis of Human Plasma Proteome using Two-dimensional Gel Electrophoresis. J Proteomics Bioinform 2: 495-499.
8. Michael K, Gorden R, Martin E, Anke S, Helmut EM, et al. (2008) Automated Calculation of Unique Peptide Sequences for Unambiguous Identification of Highly Homologous Proteins by Mass Spectrometry. J Proteomics Bioinform 1: 6-10.
9. Joanna G, Robert LJG, Raymond B, Victoria EMG, Stephen CD, et al. (2008) The Use of Conjunctival Swab for the Proteomic Characterisation of Dry Eye Syndrome. J Proteomics Bioinform 1: 17-27.
10. Annika T, Jonas F, Fredrik B, Michael N, Kaj B, et al. (2008) Proteome Analysis of Serum-Containing Conditioned Medium from Primary Astrocyte Cultures. J Proteomics Bioinform 1: 128-142.
11. Fengming G, Shufang L, Chunmei H, Guobo S, Yuhuan X, et al. (2008) The Optimized Conditions of Two Dimensional Polyacrylamide Gel Electrophoresis for Serum Proteomics. J Proteomics Bioinform 1: 250- 257.

12. Somnath T, Virendra VG, Rajat KD (2008) Pathway Modeling: New face of Graphical Probabilistic Analysis. J Proteomics Bioinform 1: 281-286.

13. Ramón A, Javier GG, Baldo O (2008) Integration and Prediction of PPI Using Multiple Resources from Public Databases. J Proteomics Bioinform 1: 166-187.

14. Fong CC, Edward AK, Mark DW, Scott JT (2008) DataBiNS-Viz: A Web-Based Tool for Visualization of Non-Synonymous SNP Data. J Proteomics Bioinform 1: 233-236.

15. Sunil K , Priya RD, Prakash CS (2008) Prediction of 3-Dimensional Structure of Cathepsin L Protein of *Rattus Norvegicus*. J Proteomics Bioinform 1: 307-314.

16. Paul K, Nathan LC, Daniel C, Clarissa D, Patrice H, et al. (2008) Global Proteomics: Pharmacodynamic Decision Making via Geometric Interpretations of Proteomic Analyses. J Proteomics Bioinform 1: 315-328.

17. Ramón A, Javier GG, Baldo O (2008) Integration and Prediction of PPI Using Multiple Resources from Public Databases. J Proteomics Bioinform 1: 166-187.

18. Sekiyama E, Matsuyama Y, Higo D, Nirasawa T, Ikegawa M, et al. (2008) Applying Magnetic Bead Separation / MALDI-TOF Mass Spectrometry to Human Tear Fluid Proteome Analysis. J Proteomics Bioinform 1: 368-373.

19. van den Toorn HWP, Mohammed S, Gouw JW, van Breukelen B, Heck AJR (2008) Targeted SCX Based Peptide Fractionation for Optimal Sequencing by Collision Induced, and Electron Ttransfer Dissociation. J Proteomics Bioinform 1: 379- 388.

20. Subramanian R, Muthurajan R, Ayyanar M (2008) Comparative Modeling and Analysis of 3-D Structure of EMV2, aLate Embryogenesis Abundant Protein of *Vigna Radiata* (Wilczek). J Proteomics Bioinform 1: 401-407.

21. Ying J, Quanjun W, Jinglan W, Songfeng W, Gang C, et al. (2008) Profiling of Phosphorylated Proteins in Human Fetal Liver. J Proteomics Bioinform 1: 437-457.

22. Gustafsson JOR, McColl SR, Hoffmann P (2008) Imaging Mass Spectrometry and Its Methodological Application to Murine Tissue. J Proteomics Bioinform 1: 458-463.

23. Thiele H, Jörg G, Peter H, Gerhard K, Martin B (2008) Managing Proteomics Data: From Generation and Data Warehousing to Central Data Repository. J Proteomics Bioinform 1: 485-507.

24. Eugenia GG, George L, Elias SM (2011) Visualizing Meta-Features in Proteomic Maps. BMC Bioinform 12: 308.

25. Casado B, Luisetti M, Iadarola P (2011) Advances in proteomic techniques for biomarker discovery in COPD. Expert Rev Clin Immunol 7: 111-123.

26. Sá-Correia I, Teixeira MC (2010) 2D electrophoresis-based expression proteomics: a microbiologist's perspective. Expert Rev Proteomics 7: 943-953.

27. Ly L, Wasinger VC (2011) Protein and peptide fractionation, enrichment and depletion: tools for the complex proteome. Proteomics 513-534.

28. Lenz T, Poot P, Gräbner O, Glinski M, Weinhold E, et al. (2010) Profiling of methyltransferases and other S-adenosyl-L-homocysteine-binding Proteins by Capture Compound Mass Spectrometry (CCMS). J Vis Exp 20: 2264.

29. Wenger CD, Phanstiel DH, Lee MV, Bailey DJ, Coon JJ (2011) COMPASS: a suite of pre- and post-search proteomics software tools for OMSSA. Proteomics 11:1064-1074.

30. Matthiesen R, Azevedo L, Amorim A, Carvalho AS (2011) Discussion on common data analysis strategies used in MS-based proteomics. Proteomics. 11: 604-619.

31. Gomes-Alves P, Neves S, Penque D (2011) Signaling pathways of proteostasis network unrevealed by proteomic approaches on the understanding of misfolded protein rescue. Methods Enzymol 491: 217-233.

32. Zeeberg BR, Liu H, Kahn AB, Ehler M, Rajapakse VN, et al. (2011) RedundancyMiner: De-replication of redundant GO categories in microarray and proteomics analysis. BMC Bioinformatics 10: 12-52.

33. Rogers S (2011) Statistical methods and models for bridging Omics data levels. Methods Mol Biol 719:133-151.

34. DeLuca DS, Marina O, Ray S, Zhang GL, Wu CJ, et al. (2011) Data processing and analysis for protein microarrays. Methods Mol Biol 723:337-347.

35. Bielow C, Gröpl C, Kohlbacher O, Reinert K (2011) Bioinformatics for qualitative and quantitative proteomics. Methods Mol Biol 719:331-349.

36. Matthiesen R, Jensen ON (2008) Analysis of mass spectrometry data in proteomics. Methods Mol Biol 453:105-122.

37. Sanders WS, Wang N, Bridges SM, Malone BM, Dandass YS, et al. (2011) The proteogenomic mapping tool. BMC Bioinformatics 22:12-115

38. Teunissen CE, Koel-Simmelink MJ, Pham TV, Knol JC, Khalil M, et al. (2011) Identification of biomarkers for diagnosis and progression of MS by MALDI-TOF mass spectrometry. Mult Scler 17:838-850.

39. Zhang Y, Li N, Brown PW, Ozer JS, Lai Y (2011) Liquid chromatography/tandem mass spectrometry based targeted proteomics quantification of P-glycoprotein in various biological samples. Rapid Commun Mass Spectrom. 25:1715-1724.

40. Sardiu ME, Washburn MP (2011) Building protein-protein interaction networks with proteomics and informatics tools. J Biol Chem 286: 23645-23651.

41. Ma ZQ, Chambers MC, Ham AJ, Cheek KL, Whitwell CW, et al. (2011) ScanRanker: Quality assessment of tandem mass spectra via sequence tagging. J Proteome Res 10:2896-2904.

42. Kolker E, Higdon R, Welch D, Bauman A, Stewart E, et al. (2011) SPIRE: Systematic protein investigative research environment. J Proteomics

43. Stunnenberg HG, Vermeulen M (2011) Towards cracking the epigenetic code using a combination of high-throughput epigenomics and quantitative mass spectrometry-based proteomics. Bioassays 33:547-551.

44. Raju K, Doulias PT, Tenopoulou M, Greene JL, Ischiropoulos H (2011) Strategies and tools to explore protein S-nitrosylation. Biochim Biophys Acta.

45. Courcelles M, Lemieux S, Voisin L, Meloche S, Thibault P (2011) ProteoConnections: a bioinformatics platform to facilitate proteome and phosphoproteome analyses. Proteomics 11:2654-2671.

46. Gauci VJ, Wright EP, Coorssen JR (2010) Quantitative proteomics: assessing the spectrum of in-gel protein detection methods. J Chem Biol 4: 3-29.

47. Haw R, Hermjakob H, D'Eustachio P, Stein L (2011) Reactome Pathway Analysis to Enrich Biological Discovery in Proteomics Datasets. Proteomics

48. McLamore ES, Porterfield DM (2011) Non-invasive tools for measuring metabolism and biophysical analyte transport: self-referencing physiological sensing. Chem Soc Rev.

49. Galan JA, Paris LL, Zhang HJ, Adler J, Geahlen RL, et al. (2011) Proteomic studies of Syk-interacting proteins using a novel amine-specific isotope tag and GFP nanotrap. J Am Soc Mass Spectrom. 22:319-328.

50. Sandro Andreotti, Gunnar W. Klau, Knut Reinert, "Antilope – A Lagrangian Relaxation Approach to the de novo Peptide Sequencing Problem," IEEE/ACM Transactions on Computational Biology and Bioinformatics, 22 Mar. 2011. IEEE computer Society Digital Library. IEEE Computer Society,

51. Medina-Aunon JA, Martinez-Bartolome S, Lopez-Garcia MA, Salazar E, Navajas R, et al. (2011) The Proteored MIAPE Web Toolkit: A User-Friendly Framework To Connect And Share Proteomics Standards. 13. Mol Cell Proteomics.

52. Knowledge-based technologies in proteomics (2011). Bioorg Khim 37: 190-198.

53. Satyavani R, Fatima A, Sundaram CS, Anabalagan C, Saritha CV, et al. (2009) Proteomic Analysis Of The "Side Population" (SP) Cells From Murine Bone Marrow. J Proteomics Bioinform 2: 398-407.

54. Narayanan A, Zhou W, Ross M, Tang J, Liotta L, et al. (2009) Discovery of Infectious Disease Biomarkers in Murine Anthrax Model Using Mass Spectrometry of the Low-Molecular-Mass Serum Proteome. J Proteomics Bioinform 2: 408-415.

55. Wang XL, Fu A, Spiro C, Lee HC (2009) Proteomic Analysis of Vascular Endothelial Cells-Effects of Laminar Shear Stress and High Glucose. J Proteomics Bioinform 2: 445-454.

56. Barh D, Misra AN (2009) Epitope Design from Transporter Targets in N. gonorrhoeae. J Proteomics Bioinform 2: 475-480.

57. Vetrivel U, Sankar P, Nagarajan NK, Subramanian G (2009) Peptidomimetics Based Inhibitor Design for HIV – 1 gp120 Attachment Protein. J Proteomics Bioinform 2: 481-484.

58. Li L, Sun C, Freeby S, Yee D, Kieffer-Jaquinod S, et al. (2009) Protein Sample Treatment with Peptide Ligand Library: Coverage and Consistency. J Proteomics Bioinform 2: 485-494.

59. Kikuta K, Tsunehiro Y, Yoshida A, Tochigi N, Hirohahsi S, et al. (2009) Proteome Expression Database of Ewing sarcoma: a segment of the Genome Medicine Database of Japan Proteomics. J Proteomics Bioinform 2: 500-504.

60. Sandra O, Henning H (2008) Standardising Proteomics Data – the work of the HUPO Proteomics Standards Initiative. J Proteomics Bioinform 1: 3-5.

61. Neha G, Sachin P, Anil P, Anil K (2008) Primer Designing for Dreb1A, A Cold Induced Gene. J Proteomics Bioinform 1: 28-35.

62. Allam AR, Kiran KR, Hanuman T (2008) Bioinformatic Analysis of Alzheimer's Disease Using Functional Protein Sequences. J Proteomics Bioinform 1: 036-042.

63. Kush A, Raghava GPS (2008) AC2DGel: Analysis and Comparison of 2D Gels. J Proteomics Bioinform 1: 43-46.

64. Zhag ZH, Hongzhan H, Amrita C, Mira J, Anatoly D, et al. (2008) Integrated Bioinformatics for Radiation-Induced Pathway Analysis from Proteomics and Microarray Data. J Proteomics Bioinform 1: 47-60.

65. Atsushi K, Yoko I, Masashi T, Yoriko T, Masao Y, et al. (2008) Development of a Data-mining System for Differential Profiling of Cell Glycoproteins Based on Lectin Microarray. J Proteomics Bioinform 1: 68-72.

66. Seenivasagan R, Kasimani R, Marimuthu P, Kalidoss R, Shanmughavel P (2008) Comparative Modeling Of Viral Protein R (Vpr) From Human Immunodeficiency Virus 1 (Hiv 1). J Proteomics Bioinform 1: 073- 076.

67. Parveen S, Asad UK (2008) Proteolytic Enzymes Database. J Proteomics Bioinform 1: 109-111.

68. Pallavi S, Vijai S, PK Seth (2008) *In Silico* Prediction of Epitopes in Virulence Proteins of *Mycobacterium Tuberculosis* H37Rv for Diagnostic and Subunit Vaccine Design. J Proteomics Bioinform 1: 143-153.

69. Vijayakumar S, Piramanaygam S (2008) SiRNA Scanner - A Fuzzy Logic Based Tool for Small Interference RNA Design. J Proteomics Bioinform 1: 154-160.

70. Rao VS, Das SK, Rao VJ, Gedela S (2008) Recent developments in life sciences research: Role of Bioinformatics. African Journal of Biotechnology 7: 495-503.

71. University of Bristol, Proteomics facility.

72. ExPASy SIB Bioinformatics Resource Portal - Proteomics Tools.html.

73. Nikolskaya A, Proteomics and Protein Bioinformatics: Functional Analysis of Protein Sequences.ppt, Protein Information Resource, Department of Biochemistry and Molecular Biology, Georgetown University Medical Center.

74. Kuznetsov A (2009) VMD: Genetic Networks Described in Stochastic Pi Machine (SPiM) Programming Language: Compositional Design. J Comput Sci Syst Biol 2: 272-282.

75. Lakshminarasimhan N, Tagore S (2009) Damages in Metabolic Pathways: Computational Approaches. J Comput Sci Syst Biol 2: 208-215.