Advances in proteomics analytical techniques

Reyhaneh Farrokhi Yekta¹, Mehdi Koushki², Nasrin Amiri DashAtan¹,*

¹Proteomics Research Center, Shahid Beheshti University of Medical Sciences
²Clinical Biochemistry Department, School of Medicine, Tehran University of Medical Sciences, Tehran, Iran

*Corresponding Author: email address: nasrinamiri91@gmail.com (N. Amiri DashAtan)

ABSTRACT

Proteins are fundamental components of cells which mediate many essential biological processes. Proteomics is a rapidly growing field for the study of proteome, the protein complement expressed by the genome of an organism or cell type. The large-scale analysis of proteins leads to a more comprehensive view of molecular and cellular pathways that improves the overall understanding of the complex processes supporting the living systems. The analysis of proteome is significantly challenging due to high dynamic range and difficulties in assessment of low abundance proteins and the absence of efficient purification and identification techniques. A variety of methods have been utilized for protein studies including gel-based techniques, protein microarrays, mass spectrometry-based approaches such as MALDI and SELDI, high and ultra-performance liquid chromatography and fourier transform ion cyclotron resonance mass spectrometry. NMR spectroscopy and X-Ray crystallography methods are also used for structural study of proteins. This review aims to give a brief overview of some of the above techniques and their most recent advances. We also introduce Proteominer, a recent protein enrichment technology for the exploration of the entire proteome content.

Keywords: Mass Spectrometry; Electrophoresis; Proteomics Techniques.

INTRODUCTION

The analysis of proteins has undergone a major revolution over the past 20 years from amino acid analysis and Edman sequencing to today mass spectrometry platforms[1]. One of the first techniques used in proteomics is 2D-gel electrophoresis. The great advantage of modern 2D gel-based proteomic experiments is that it is simple and robust technology and can be quickly implemented into most laboratories[2]. Unlike gel-based approaches, MS-based relative quantitation techniques usually rely on an initial digestion of the protein, and the subsequent protein quantification is actually based on the quantitation of proteotypic peptides that act as surrogates for the proteins of interest[3]. Mass spectrometry is an extraordinary information rich technology that is capable of detecting tens of thousands of peptides generated in a single separation[4]. Mass spectrometry-based approaches are uniquely well suited in terms of throughput and sensitivity to handle proteome-wide investigations[5]. There are two main types of mass spectrometric analysis approaches, “Top-Down” and “Bottom-Up” or shotgun. In “Bottom-Up” proteomics, peptides generated from enzymatic proteolysis of proteins are analyzed in a mass spectrometer[1]. In top-down proteomics, 100% sequence coverage is obtained and PTM combinations are preserved leading to precise identification of proteins. In shotgun proteomics, protein mixtures are proteolytically digested before tandem mass spectrometry (MS/MS) analysis. To reduce sample complexity and increase the chances of identifying low abundant proteins, fractionation techniques are performed such as multidimensional protein identification technology approach (MudPIT), where peptides are separated by strong cation exchange (SCX) and reversed phase chromatography prior to their identification by mass spectrometry (MS). Protein and peptide fractionations using electrophoresis is also common because of its high capability and resolving power[6]. Accurate quantification of
proteins at proteome level has become one of the key issues in protein science. Therefore, quantitative proteomics, emerged as a new research area in the last decade. Stable isotope labeling and label free techniques are used for this purpose. Label free methods include ion intensities of peptides and spectral counting but its usage is limited because of high variability in sample preparation and instrumental analysis. This challenge is addressed by stable isotope labelling methods. Since such labeling does not affect the chemical properties of proteins, the heavily and lightly labeled samples could be co-eluted from the LC-column, followed by the simultaneous analysis in the mass spectrometer[7]. Some of the major isotope labeling techniques are SILAC, iTRAQ and ICAT. Structural proteomics is the systematic investigation of the three-dimensional structures of the protein products of genes. Because of the challenge in structure determination of large numbers of proteins, the field of structural proteomics developed [8]. NMR spectroscopy and X-Ray Crystallography are widely used in this area. The following paper aims to discuss briefly major technologies in the field of proteomics. MS-based methods, two-dimensional gel electrophoresis and some novel gel- and non-gel-based methods, isotope-labeling and label-free techniques are studied. After talking about microarray and proteominer technologies, structural proteomics methods will be discussed.

MASS-BASED METHODS

MALDI-TOF-MS

MALDI (Matrix-Assisted Laser Desorption/Ionization), in combination with TOF-MS (time of flight- mass spectrometry), has emerged as a valuable technique for identification of proteins. MALDI can be applied for compounds in the m/z range of 500 to over 100,000[9]. MALDI-TOF instruments are relatively simple to use, have high mass accuracy and are reasonably tolerant of contaminants and solvents[10]. The first reports of MALDI-TOF-MS biochemical analysis were published in the late 1980s from Karas and Hillenkamp lab[11]. MALDI has significantly revolutionized approaches to the study of biomolecules. MALDI is initiated by mixing the sample solution with matrix material and depositing the mixture on a specially designed MALDI sample target. After evaporation of the solvent, the sample–matrix crystals are irradiated using laser beam of high irradiance and short pulse widths to simultaneously desorb and ionize the sample and matrix molecules into the gas phase (see Figure1).

An essential key to success of MALDI is a matrix that is able to absorb a large amount of energy at the wavelength of the laser radiation, and then relays it to the sample molecules in a controlled manner to permit desorption of even massive molecules as intact gas-phase ions. The MALDI-generated ions are mainly singly protonated molecules. Oligomeric ions and doubly and triply charged protonated ions are also formed. Because the irradiating laser beam is pulsed, MALDI is optimally combined with a TOF mass analyzer. The unlimited mass range of TOF and its ability to acquire the entire spectrum from a single laser pulse event are other factors in favor of the MALDI/TOF-MS combination. A TOF analyzer measures the time taken for the gas-phase ions to travel from the ionization source to the detector, which is then related to the m/z ratio[12-14]. TOF analyzer is based on the fact that ions with the same energy but different masses travel with different velocities. Basically, ions formed by a short ionization event are accelerated by an electrostatic field to a common energy and travel over a drift path to the detector. The lighter ones arrive before the heavier ones and a mass spectrum is recorded. Measuring the flight time for each ion allows the determination of its mass. MALDI –TOF has become a well-known acronym for many researchers. A variety of laser systems have found applications in MALDI analysis, and the most common ones use UV lasers such as the N₂ laser (337 nm).
IR lasers have also been used to produce the MALDI effect. UV and IR lasers yield similar spectra for proteins, although better resolution has been obtained for some proteins with an IR laser\cite{15}.

**SELDI Technique**

SELDI-TOF-MS is an adaptation of MALDI-TOF-MS using surface-modified target plates\cite{16}. This technology was introduced in 1993 by Hutchens and Yip for the first time. SELDI is an advanced approach to protein profiling and new biomarker discovery. SELDI-TOF has developed in clinical research world, especially because of its high-throughput capability. Main features are its sensitivity of detection, accuracy of quantification and its capability of generating reproducible patterns in different laboratories\cite{17}. Since, biological samples (like blood and urine) are complex mixtures, SELDI is a powerful tool that overcomes purification and separation of proteins prior to mass spectrometry analysis (figure 2). In this technique, microliters of the sample are incubated onto the chip surface and the chips are then washed using washing buffers. The sample is typically analyzed with time-of-flight mass spectrometry\cite{18}.

The differential expression data obtained from this technology has been used for identification of biomarker candidates for various cancer types, such as prostate\cite{19}, pancreas\cite{20}, lung\cite{21}, breast\cite{22}, melanoma\cite{23}, and liver cancers\cite{24}. For most cancers, survival rates depend on the early detection of the disease. Novel mass spectrometry (MS)-based technologies in particular, SELDI-TOF-MS, have brought the hope of discovering new cancer-specific biomarkers in biological samples and have shown promising results in the recent literature\cite{25}.

The SELDI-TOF-MS technology is not only able to find single protein biomarkers but is also able to identify biomarker expression patterns. Proteomic pattern analysis is a novel approach for the diagnosis of diseases\cite{26}.

In summary, there is considerable hope that this new proteomic technology will be used significantly to screening-test development in routine clinical practice\cite{25}.

An advantage of SELDI-TOF-MS is its relatively high tolerance for salts and other impurities. The sample requirement is low and sample volume can be freely chosen from 0.5l up to around 400µl\cite{26}.

---

**FT-ICR mass spectrometry**

Fourier transform ion cyclotron resonance mass spectrometry, also known as Fourier transform mass spectrometry, is a type of mass analyzer (or mass spectrometer) for determining the mass-to-charge ratios (m/z) of ions based on the cyclotron frequency of the ions in a fixed magnetic field. The FT-ICR mass analyzer, introduced in 1974, has the highest mass resolving power and best mass measurement accuracy among current mass analyzers. The orbitrap, another Fourier transform mass analyzer, invented in 1999, has been widely distributed since its commercial introduction in 2004. In an FT-ICR instrument, ions are first generated at the source (ESI, APCI, APPI, or MALDI), and then injected into an ion trap mass analyzer cell in the center of a magnetic field (The highest field of FT-ICR instrument is currently 15 T, and 21 T systems are under construction). Fourier transform ion cyclotron resonance mass
spectrometry is based on image current detection of coherently excited ion cyclotron motion. It is now one of the most sensitive methods of ion detection in existence and has almost unlimited resolution \( >10^7 \), with most experiments taking place in the \( 10^5 \) to \( 10^6 \) range. FT-ICR-MS is also a powerful tool for conducting ion–molecule reactions and for structure elucidation studies. Because of these useful features, FT–ICR–MS in conjunction with ESI has emerged as the most powerful form of mass spectrometry for the analysis of biomolecules[27-29].

GEL-BASED METHODS

Two-dimensional gel electrophoresis

Two-dimensional gel electrophoresis (2DE) has become the most widely used separation tool in proteomic analysis. 2DE is especially useful in expression proteomics, where comparative analysis of the expression of proteins exposed to environmental factors and those physically undisturbed is the purpose of research[30]. This method was first described by O’farrell in 1975 and has developed till then. 2D electrophoresis consists of two tandem steps (dimensions). The first dimension is called isoelectric focusing (IEF). In this step, proteins are separated based on their isoelectric points in an electric field in a pH gradient. The gradient is generated using both carrier ampholytes or immobilized pH gradient (IPG) strips. IPG is obtained by copolymerization of acrylamide with immobilin. The pH gradients with IPG are more stable and reproducible. Today, commercially IPG strips are available with different length and pH ranges. Second dimension in 2D electrophoresis is SDS-PAGE (sodium dodecylsulfate polyacrylamide gel electrophoresis) which separates the proteins according to their molecular weights. After separation of the proteins on the gel, they are visualized by staining with different dyes usually coomassie blue, silver nitrate, and fluorescent dyes. The stained spots are then excised and digested for further identification.

2D-DIGE

2D-DIGE (2-dimensional difference gel electrophoresis) technique was first described by Jon Minden’s laboratory. It relies on pre-electrophoretic labeling of samples with one of fluorescent CyDyes (Cy2, Cy3 and Cy5) allowing multiplexing of samples into the same gel (figure 3). 2D-DIGE is effectively used for the study of various systems and enables detection of subtle changes in protein expression than conventional 2D-PAGE[31]. 2D-DIGE has some advantages. It is labor and time saving and produces accurate and reliable results. Furthermore, loss of proteins is reduced because no post-electrophoretic processing is needed.

Tube Gel Electrophoresis

Tube gel electrophoresis utilizes a tube gel column to separate proteins which are then collected as they elute from the end of the gel column. The use of tube gel electrophoresis was further expanded with the invention of gel-eluted liquid fraction entrapment electrophoresis (GELFrEE). It uses a sample collection chamber in which fractions are manually collected which ensures that higher molecular weight proteins are not continually diluted and dispersed across many fractions. It has a short gel column which reduces separation time 75%[5], (Figure 4).
Off-gel Electrophoresis

Gel free techniques are very important in separating proteins. A gel-free approach in protein separations is immobilized pH gradient (IPG) IEF where peptides can be recovered from the liquid phase (offgel electrophoresis) which leads to efficient protein fractionation and identification [32]. Jonson and Rilbedeveloped a gel-free multicompartmentelectrolyser electrophoretic device with IPG technology which provided better pI resolutions. A further technology is developed by Girault et al. which includes adapting the off-gel IEF to a multiwell format. The multiwell device is composed of different compartments of 100 or 300 μL open at the top and bottom extremities and placed on an IPG gel conditioned with a thin layer of solution containing buffers. There is no fluidic connection between the wells, and the charged proteins migrate through the gel across the wells under the applied electric field until they reach net charge of zero (pI), in which the proteins will be in the solution and can be recovered in the liquid phase[32].

STABLE ISOTOPE LABELLING METHODS

In these methods, proteome samples are labeled and then mixed and subjected to LC/MS analysis. The quantification of proteins is achieved by comparing of MS peak intensities from labeled and non-labeled samples. A variety of isotopic labeling techniques can be used for relative quantification, including SILAC, iTRAQ and ICAT.

Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC), is a metabolic labeling strategy that encodes whole cellular proteomes. Cells are grown in a culture medium where the natural form of an amino acid is replaced with a stable isotope form such as arginine with six 13C atoms. Incorporation of the “heavy” amino acid occurs through cell growth and protein synthesis. The “light” and “heavy” proteomes belonging to two samples are then distinguished via mass spectrometry[33].

“SILAC” has some potential limitations too. For example, although metabolic labeling is an effective way to uniformly incorporate isotopic tags into proteins, in practice, it is not always feasible as seen with clinical samples and some model organisms. Also the ion intensity for each peptide is distributed between several isotopic peaks which lowers the total number of peptide identifications from a given sample[34].

iTRAQ (isobaric tags for relative and absolute quantitation) reagents are a set of multiplexed amine specific stable isotope reagents which consists of a reporter group based on N,N-dimethylpiperazine, a mass balance carbonyl group, and a peptide-group. When iTRAQ reagent reacts with a peptide, it forms an amine linkage to any peptide amine like lysine amino group. A new method known as “NeuCode SILAC” is recently introduced which is a combination of SILAC metabolic labeling with multiple isobaric tags. This method has remarkable efficiency and
resolution to distinguish between isotopes of similar mass, based up on differences in nuclear binding energy[34].

The ICAT peptide labeling technique differentiates between two populations of proteins using reactive probes that differ in isotope composition. ICAT reagents consist of a protein-reactive group, a linker region and a biotin tag. The two different isotope tags are generated by using linkers composed of either eight deuterium atoms (d8, heavy reagent) or eight hydrogen atoms (d0, light reagent). A reduced protein sample from one specimen is derivatized with the isotopically heavy version of the ICAT reagent, while the other reduced protein sample is derivatized with the isotopically light version of the ICAT reagent. The two samples are combined and digested with a protease, such as trypsin or Lys-C, to produce peptide fragments. The combined sample is then subjected to avidin affinity chromatography and only cysteine-containing peptides are thus retrieved. MS is used to reveal the ratio of the isotopic molecular weight peaks that differ by 8 Da, and this gives a measure of the relative amounts of each protein from the original samples[35].

LABEL FREE METHODS

Label-free approaches are divided into two main categories: spectral counting and mass spectrometric signal intensities. Spectral counting implies a counting and comparison of the number of fragment ion spectra (MS/MS) for peptides of a given protein. Due to the role of protein concentration on the number of tandem mass spectra of a peptide, a relative quantification of proteins between different samples is possible. Signal intensity method, relies on the intensity of the mono-isotopic mass peak and measurement of chromatographic peak areas[36].

PROTEIN MICROARRAY

Novel proteomics technology is valuable for studying of whole proteome and network analysis. Protein microarrays, an emerging class of proteomic technologies, are fast becoming critical tools in biochemistry and molecular biology[37]. Protein microarray is a high throughput tool for studying the biochemical activities of proteins, tracking their interactions, and determining their function on a large scale[38]. Protein microarray chips that contain immobilized proteome, are being developed to simultaneously analyze protein function and protein-protein, protein-DNA or protein-ligand interactions in a high-throughput in vitro manner in a single experiment[39]. The chip usually consists of a support surface such as a glass slide, nitrocellulose membrane, bead, or microtiter plate to which an array of capture proteins is bound[18]. This method can be divided to two types, forward-phase and reverse-phase arrays[40, 41]. In forward-phase protein arrays (FPPA), an antibody as a capture molecule for a target protein is immobilized onto a glass slide robotically and cell lysate that contains target proteins is incubated onto the slide, finally bounded proteins are detected using secondary labeled antibody[40, 42]. In this type, many target proteins in a sample can be identified simultaneously. In reverse-phase protein arrays (RPPA), protein mixture is immobilized onto glass slide and probed with a specific antibody against a protein of interest. RPPA assays are commonly used in tissue microarray and cell and tissue lysate microarray, furthermore proteins of interest in a complex sample can be detected[43]. New protein microarray platforms such as self-assembling arrays are emerging, which promise a much easier and wider use of the technology to probe protein interaction and function[44].

PROTEOMINER: PROTEIN ENRICHMENT TECHNOLOGY

Proteominer protein enrichment technology is a novel and simple sample preparation tool used to compress the dynamic range of protein concentrations in complex biological samples. For example, albumin and IgG in serum or plasma make the detection of medium and low-abundance proteins extremely challenging. Proteominer technology provides a method for overcoming this challenge allowing the exploration of the entire proteome. Proteominer technology is based on the interaction of complex protein samples with a large, highly diverse library of hexapeptides bound to chromatographic supports. In theory, each unique hexapeptide
binds to a unique protein sequence. Because the bead capacity limits binding capacity, high-abundance proteins quickly saturate their ligands and excess protein is washed out during procedure. In contrast, low-abundance proteins are concentrated on their specific ligands, thereby decreasing the dynamic range of proteins in the sample. When analyzed in downstream applications, the number of proteins detected is dramatically increased[45], (Figure 5).

**STRUCTURAL PROTEOMICS**

**X-Ray Crystallography**

X-Ray crystallography is one of the two major methods for the elucidation of protein structures besides NMR spectroscopy. In crystallography, an X-ray beam is diffracted by a protein crystal, which is a regular lattice of protein molecules arranged in a repeating pattern and held together by non-covalent forces. Single protein molecules cannot scatter sufficient X-rays to be detected, but X-rays scattered from a protein lattice combine in intensity and can be recorded. Processing of these intensities yields a model of the density of electrons in the protein crystal. Obtaining well diffracting protein crystals often constitutes a major bottleneck, since many proteins are difficult to crystallize or sometimes do not crystallize at all. If crystallization is successful, the use of X-ray crystallography is advantageous, since it offers structural information at atomic resolution without a size limit. The X-ray electron density map shows a finely detailed, albeit static, "snapshot" of the protein, typically in its lowest-energy conformation[46].

**NMR Spectroscopy**

NMR spectroscopy plays a major role in determination of proteins 3D structures. It is utilized to determine structures of smaller proteins that fail to form crystals suitable for structure determination by x-ray crystallography, to screen structural candidates for folding and aggregation state, and to screen proteins for binding of metal ions, cofactors, or their small molecules. Protein NMR spectroscopy requires 1 to 5 mg of purified protein, and the protein must be labeled with stable isotopes (like nitrogen-15 or carbon-13). 1D H1 NMR and 2D 15N-1H HSQC spectrums usually provide reliable results. Some other platforms that can be used include 3D 15N-1H NOESY-HSQC, 2D-1H-13C HSQC, 3D-HNCO, 3D-HNCA, 3D-CCONH, 3D-HCCH-TOCSY and so on[8]. NMR spectroscopy has also been used for “in-cell” studies of proteins because it non-invasively gathers data from the cells. It is an ideal tool for gaining information about protein dynamics at the atomic level[47].This technique has been successfully applied in Escherichia coli, Xenopus laevis oocytes and HeLa host cells. 2D-1H-15N (or 1H-13C) HSQC platform is usually used. However, the technique requires that the protein of interest is expressed to intracellular concentrations sufficient for NMR detection that are greater than concentration of most cellular proteins[48]. In-cell NMR has several obstacles;
Few proteins provide high quality NMR spectra inside cells.
Second, in cell NMR is limited by the life span of the cells in the NMR tube and even if cells do not lyse, they may leak the target protein into the media.[47]. The majority of protein NMR studies are carried out in the liquid state. Although solution NMR methods have been used to characterize unfolded states of proteins, solid-state NMR techniques provide both qualitative and quantitative structural information about protein folding, including detailed nature of conformational distributions in partially folded and unfolded states at equilibrium and the time dependence of structural distributions after sudden changes in solvent conditions.[49].

CONCLUSION
Large scale study of proteins, as a growing field in biological sciences, has gained much attraction in recent years. One of the most important factors in success of the field is the evolution of novel techniques for the separation and identification of proteins. There are gel-based, such as 2D-PAGE and 2D-DIGE, and non-gel-based methods. Mass based methods such as MALDI are the most utilized techniques in identification of proteins. Stable isotope labeling, such as SILAC and iTRAQ, and label free methods are used for the accurate quantification of proteins at proteome level. In the field of structural proteomics, NMR spectrometry and X-Ray crystallography methods are used.

REFERENCES
17. Semmes OJ FZ, Adam BL et al. Evaluation of serum protein profiling by surface enhanced laser desorption/ionization time-of-flight mass
41. Liotta LA EV, Mehta AI et al. Protein
47. Conggang Li ML. Protein dynamics in living cells studied by in-cell NMR spectroscopy. FEBS Letters 2013.
48. Shekhtman A. Protein NMR techniques. 3 ed2012.