Proteomics Technology

DELFINA C DOMINGUEZ, ROSANA LOPES, M LORRAINE TORRES

Proteomics techniques are essential tools for protein detection and characterization. Besides several advances in the proteomics field, the two-dimensional electrophoresis (2-DE) technique is the most important method for protein separation. The combination of 2-DE technique, new advances in mass spectrometry and bioinformatics promises to unveil protein function and pathological mechanisms of disease.

ABREVIATIONS: 2-DE = two-dimensional electrophoresis; CHAPS = (3-[(3-Cholamidopropyl)-Dimethylammonio]-1-Propane Sulfonate); ESI - MS= electrospray ionization mass spectrometry; ICAT = isotope-code affinity tag; IPG = immobilized pH gradient; LC-MS = liquid chromatography mass spectrometry; MALDI = matrix assisted laser desorption ionization; MALDI-TOF = matrix assisted laser desorption ionization time-of-flight; mRNA = messenger ribonucleic acid; pI = isoelectric point; SELDI-MS = surface-enhanced laser desorption/ionization time-of-flight.

INDEX TERMS: ICAT; LC-MS, MALDI; protein expression; proteomics; SELDI-MS; two-dimensional electrophoresis.

Clin Lab Sci 2007;20(4):239

***Delfina C Domínguez PhD** is **associate professor**, **Rosana Lopes PhD** is **assistant research professor**, and **M Lorraine Torres MS** is **faculty and program director**, all of the Clinical Laboratory Science Program, The University of Texas at El Paso, College of Health Sciences, El Paso TX.*

The Focus section seeks to publish relevant and timely continuing education for clinical laboratory practitioners. Section editors, topics, and authors are selected in advance to cover current areas of interest in each discipline. Readers can obtain continuing education credit (CE) through P.A.C.E.® by completing the continuing education registration form, recording answers to the examination, and mailing a photocopy of it with the appropriate fee to the address designated on the form. Suggestions for future Focus topics and authors, and manuscripts appropriate for CE credit are encouraged. Direct all inquiries to the Clin Lab Sci Editorial Office, IC Ink, 858 Saint Annes Drive, Iowa City IA 52245. (319) 354-3861, (319) 338-1016 (fax). ic.ink@mchsi.com

***Address for correspondence:** Delfina C Domínguez PhD, University of Texas at El Paso, Clinical Laboratory Science, College of Health Sciences.1101 N. Campbell Street. El Paso TX, 79902-4238. (915) 747-7238, (915) 747-7207 (fax). Delfina@utep.edu.*

***Delfina C Domínguez PhD** is the **Focus: Proteomics** guest editor.*

LEARNING OBJECTIVES

1. Recognize the different tools applied to proteomics analysis.
2. Differentiate the procedures utilized in proteomics studies starting from protein sample preparation to protein identification methods.
3. Identify the steps needed for analysis of protein expression by software packages.
4. Distinguish the methods employed for qualitative and quantitative protein identification.

After geneticists sequenced the human genome and learned of the poor correlation between mRNA and protein numbers, special attention was given to the gene products, the proteins. From approximately 25,000 predicted genes in human beings,[1] it is expected that there will be more than 500,000 protein products as a result of splicing and protein modifications.[2,3] Proteomics analyses utilize several techniques for protein separation, detection, and identification. For separation of simple and complex protein mixtures, two-dimensional electrophoresis (2-DE) is the technique of choice. Several staining procedures can be applied to 2-DE gels for protein detection. Mass spectrometry is considered the best method for protein identification.

TWO-DIMENSIONAL ELECTROPHORESIS (2-DE)

Sample preparation

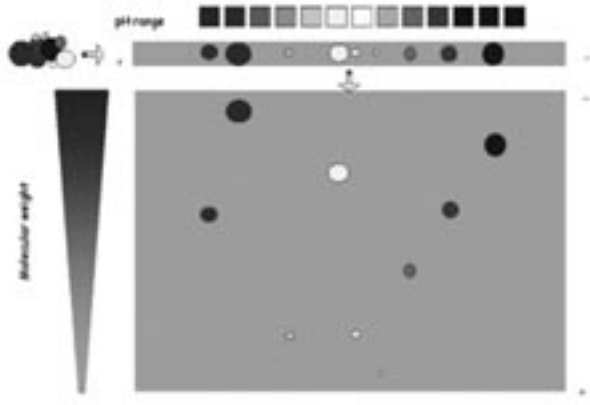
A wide variety of samples can be utilized to identify and characterize proteins in proteomic studies. Some of these samples include: samples fractionated from an organism (eukaryote or prokaryote), tissue, cell lysates, and physiological fluids.[2,4] A careful research design should be planned to acquire a meaningful representation of the proteins of interest. It is essential to ensure that proteins are soluble in order to

obtain an accurate protein analysis. Protein solubilization is achieved utilizing chaotrops (urea), reducing agents (dithiothreitol), detergents (CHAPS, Triton-X), buffers (Tris), and <0.2% ampholytes. In addition, size, charge, and isoelectric point of proteins of interest are important characteristics for sample preparation. Moreover, to reduce complexity of protein data, subcellular fractionation methods can be used to analyze proteins from specific organelles such as endoplasmic reticulum, mitochondria, nucleus or cell wall and membrane fractions from microorganisms.[3-5] Some additional methods such as selective fractionation and chromatography can also be utilized to obtain a specific set of proteins from a complex mixture.[2,6,7]. A good sample preparation prevents protein aggregation and enzymatic and chemical degradation. It also eliminates or digests nucleic acids producing a standardized protein sample. The Bradford and Lowry methods are commonly used for protein concentration measurements.

2-DE - first dimension

The first step toward protein separation is to determine its isoelectric point. The isoelectric focusing of proteins can be determined using immobilized pH gradients or ampholytes. In general, the pH gradient used for protein separation is in the range of pH 3-12. To obtain a reference map displaying the majority of proteins, a broad-range of pH gradient (pH 3-10) should be employed, but if a special category of proteins with a specific pH range is of interest, narrow pH gradients will increase the gel resolution and improve the detection of low abundance proteins. Nonlinear pH gradients are used to improve visualization of proteins in the middle of the pH range. The estimation of the isoelectric point of a protein spot can be obtained through the position of a known protein standard present in the same pH gradient. Rehydration of immobilized pH gradient (IPG) requires a minimum of 11 hours. During active rehydration a low voltage is applied while the protein sample is absorbed by the IPG. The electric current helps the protein absorption into the IPG but there is possibility of loss of low abundance proteins during this process.[3] In contrast, during passive rehydration the protein sample is absorbed into IPG without presence of an electric field. Sample application by cup loading should be avoided due to the tendency of artifacts production.[8] The voltage conditions and resolution of proteins for first dimension are dependent on the pH gradient, length of IPG strip, and electrical field applied (Figure 1).

Figure 1. Isoelectric focusing



Isoelectric focusing is the first step in the separation of proteins by 2-DE. Proteins move through a medium in an electrical field until their net charge is zero. The pH at which the protein stops migrating in an electric field is called the isoelectric point (pI) and it is determined by the charged groups in the protein. For every protein there is a specific pH at which its net charge is zero.[23]

2-DE – second dimension

In the second dimension, all proteins separated in the first dimension will be resolved according to their molecular weight. Different percentages of acrylamide determine the pore size of gels. The size of the pores created in the gel is inversely related to the amount of acrylamide used. For example, an 8% polyacrylamide gel will have larger pores in the gel than a 15% polyacrylamide gel. Gels with a low percentage of acrylamide are typically used to resolve large proteins and high percentage gels are used to resolve small proteins

Gradient gels have different concentrations of acrylamide allowing better resolution of proteins. Gradient gels are cast with a higher concentration of acrylamide at the bottom of the gel and a low concentration of acrylamide at the top of the gel. The applications of gradient gels include the determination of protein molecular weights and the separation of molecules which co-migrate on uniform gels.

Single percentage acrylamide gels will favor the separation of proteins in a narrow pH range. The run time for second dimension depends on size of gel and may range from 30 minutes for mini-gels to six hours for larger gels.

Protein detection in gel

The staining choice for 2-DE gels depends on the protein sample characteristics. For example, if low-abundance proteins are targeted, 0.1-1mg/mL of protein sample should be loaded, and a very sensitive staining should be employed

such as silver or fluorescent dyes.[9] Double staining can be performed with Coomassie brilliant blue and silver or Coomassie brilliant blue and fluorescent dyes. This procedure allows investigation of protein staining sensitivity throughout the gel once proteins interact differently with diverse types of staining.[9]

IMAGING AND PROTEIN ANALYSIS METHODS

A variety of equipment can be utilized for acquisition of 2-DE gel imaging such as cameras, scanners, and densitometers. Different software packages are available for analysis of protein expression such as PDquest (Bio-Rad), ImageMaster 2D (GE Healthcare), Progenesis Workstation (Nonlinear Dynamics), and Melanie (Geneva Bioinformatics).

A classical 2-DE analysis is based on five features: 1) Normalization,[10] 2) protein spot detection, 3) generation of a matched set of images, 4) relative quantitation of protein spots (Figure 2), and 5) statistical analysis of data. Several web sites are available for matching peptides to known protein sequences such as MASCOT (http://www.matrixscience.com/),[11] PepMAPPER (http://www.molecularstation.com/bioinformatics/link/detail/link-453.html),[12] Aldente Peptide Mass Fingerprinting (http://ca.expasy.org/tools/aldente/),[13] and ProteinProspector (http://prospector.ucsf.edu/).[14]

PROTEIN IDENTIFICATION PROCEDURES

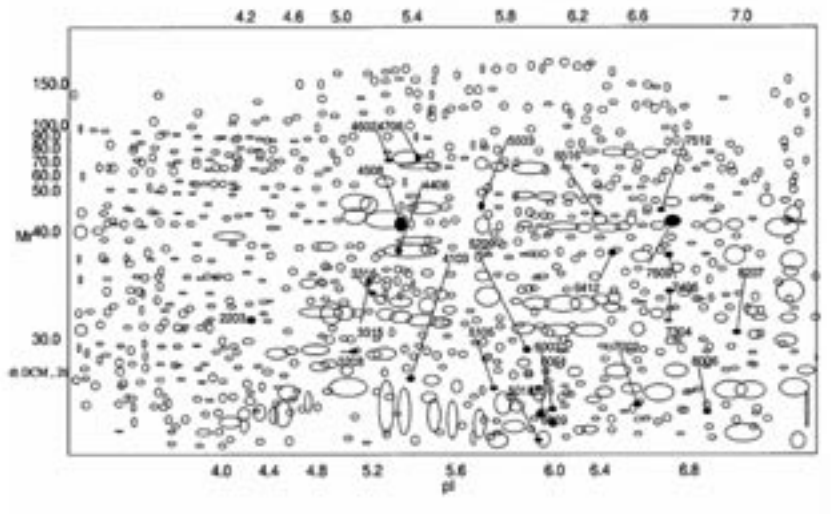
Qualitative mass spectrometry analysis

Advances in instrumentation have made mass spectrometers an essential tool for protein identification in proteomic analyses. Often, after 2-DE the protein spot is excised and subjected to in-gel trypsin digestion. The fragmented peptides are identified by mass spectrometry. Trypsin cleaves proteins on the C-terminal side of lysine and arginine residues. The mass spectra data of peptides identified are matched to known protein database sequences.[15]

In mass spectrometry (MS) the protein molecules are ionized and subsequently separated according to their mass-to-charge ratio. The separated ions are then detected by the analyzer.[15] For many years MS was used only for small and heat stable compounds due to the difficulty in ionizing the molecules effectively and the inability to transfer the ions from the condensed phase into the gas phase without excessive fragmentation.[16]

Two techniques that revolutionized MS were developed during the 1980s. These techniques were matrix assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI).[15,16,17] MALDI is known as "soft ionization" and utilizes a matrix, alpha-cyano-4-hydroxycinnamic acid and 2, 5 dihydroxybenzoic acid where peptides are embedded. Proteins to be analyzed are vaporized by a pulse-UV laser beam. Ionized proteins are accelerated in an electrical field and their mass-to-charge ratio measured.[15] This is the key feature of this ionization method.[15,17] This method is widely used to identify tryptic peptides from simple protein mixtures because sample identification can be done in less than two hours depending on the number of samples, it is sensitive to the femtomole range and inexpensive. The most common MALDI MS is MALDI-TOF (time-of-flight) (Figure 3). The mass-to-charge ratio in TOF analyzers is deduced from the flight time the ionized protein requires to travel through a tube of specific length under vacuum.[16] TOF analyzers give high mass accuracy and resolution.[18] Another variety of MALDI is the MALDI-QqTOF.

Figure 2. Relative quantitation of mitochondrial proteins map



Map shows position and standard spot numbers (SSP) of quantitation analysis from PDQuest (version 7.1) software. Position of proteins spots down-regulated by 2.0 fold or greater are shown in black. Mr = molecular weight; pI = isoelectric point.[24]

In this special case a MALDI-TOF spectrometer is connected to a quadrupole, which guides the ions to the TOF analyzer where the mass analysis takes place.[16]

If the identity of a protein cannot be obtained by MALDI-TOF from known protein databases, the electrospray ionization-MS/MS (ESI-tandem MS) method should be employed. In ESI-MS/MS, the protein sample is in liquid phase. The sample is pumped through a hypodermic needle at high voltage. The ionized molecules are then analyzed by the mass spectrometer (Figure 4).[18] ESI-MS/MS offers the advantage of providing the peptide molecular weight plus sequence information. Therefore, it is much easier to identify the protein.[15] Still, ESI-MS/MS has some disadvantages compared to MALDI. Sample conditioning plays an important role.[17] ESI-MS/MS analysis time is limited since the liquid sample is consumed during the measurements. Also ESI-MS/MS is concentration-sensitive and requires desalted samples. Further, one to five samples may cost $450.00. The presence of detergents, metals, and salts may cause problems and signal suppression.[15,17] These problems can be overcome by combining liquid chromatography (LC) with MS analysis. During LC-MS the protein sample is eluted from a chromatographic column and then analyzed by mass spectrometry. This type of analysis is especially useful when complex mixtures are analyzed.[17]

Quantitative mass spectrometry analysis

To quantify a protein sample, stable isotopes are employed to label digested protein. The isotope-coded affinity tagging (ICAT) developed by Gygi and others (1999) is able to simultaneously analyze and accurately identify and quantitate proteins from complex mixtures.[19] It is possible to compare the proteins present in two proteomes. For example, the protein samples are mixed with the ICAT reagents, which consist of a cysteine reactive moiety, a biotin affinity tag for isolation and purification, and a linker region consisting of hydrogens (light tag) or deuteriums (heavy tag).[19,20] The protein sample then is subjected to tryptic digestion followed by LC-MS/MS. Proteins are quantified based on signal intensities of identical peptides labeled with either the heavy or the light tag form the ICAT reagent. This method has the advantage of omitting the 2-DE separation step.[19] However, the main limitation of the ICAT method is that it requires the labeling of a free cysteine residue on the protein. Therefore, proteins lacking a free cysteine residue will not be identified.[5,17] This method is a low throughput method in which only two samples can be done per run.[21]

Figure 3. MALDI-TOF mass spectrometer

The MALDI plate showing protein sample precipitated in a matrix (alpha-cyano-4-hydroxycynnamic acid) on the left (3a) and mass spectrum obtained from the analysis of tryptic peptides performed in a Voyager-DE STR (3b).[24]

Another quantitative MS method is the surface-enhanced laser desorption/ionization time-of-flight (SELDI-TOF) mass spectrometry, which has a quantitative protein chip and it is known as SELDI-TOF-MS-based Protein Chip®. The principle of this method is very similar to the MALDI. In both methods the peptides are cocrystalized on a matrix or surface and peptides are vaporized by pulsed-UV laser beam. The ionized proteins then are accelerated in an electrical field and the mass-to-charge ratio is deduced from their velocity.[22] In SELDI protein solutions are applied to a protein chip array that has a specific chromatographic surface. The proteins interact with the array surface and are selectively bound according to their surface interaction. Once the proteins are bound, the surface is washed with a buffer solution, ensuring sample cleaning. Subsequently a laser beam causes desorption and ionization of proteins. A specific laser beam scans the entire protein chip and performs repeated readings of a single spot. The laser never reads the same position twice. An average spectrum is obtained and good correlation between the signal intensity and concentration of the analyte is obtained for the different protein samples.[21,22] The SELDI-TOF ProteinChip® system is limited to bound proteins and is unsuitable for analysis of proteins >100 kDa molecular weight.[21]

CONCLUSIONS

2-DE and MS analyses are crucial technologies in the study of proteomics. Great advances in instrumentation developed in the last few years have made it possible to rapidly and accurately analyze complex protein mixtures. A wide variety of systems have been created including quantitative MS.

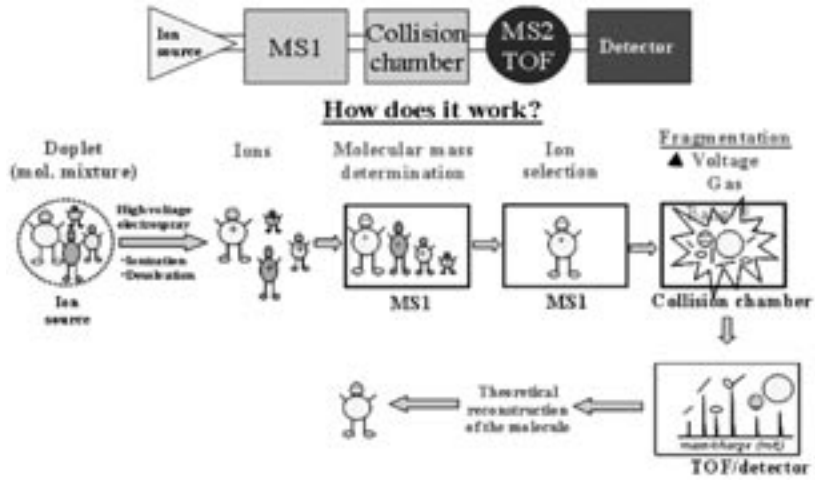
Although the accelerated advancement in MS technology has been driven by protein research, it can be anticipated that MS will also be used for the analysis of other molecules such as lipids and carbohydrates. Specific challenges still remain in the areas of protein quantitation and software development for data acquisition, data interpretation, and presentation. Further improvement in MS technology will make MS an even more powerful tool for proteomic analysis.

Clin Lab Sci encourages readers to respond with thoughts, questions, or comments regarding this article. Email responses to ic.ink@mchsi.com. In the subject line, please type "CLIN LAB SCI 20(4) FO DOMINGUEZ". Selected responses will appear in the Dialogue and Discussion section in a future issue. Responses may be edited for length and clarity. We look forward to hearing from you.

REFERENCES

1. Taylor S, Fahy WE, Ghosh SS. Global organellar proteomics. Trends Biotechnol 2003; 21:82-8.
2. Righetti PG, Castagna A, Antonioli P, Boschetti E. Prefractionation techniques in proteome analysis: the mining tools of the third millennium. Electrophoresis 2005; 26:297-319.
3. Sanchez JC, Rouge V, Pisteur M. Improved and simplified in-gel sample application using reswelling of dry immobilized pH gradients. Electrophoresis 1997; 18:324-7.
4. Huber LA, Pfaller K, Vietor I. Organelle proteomics: implications for subcellular fractionation in proteomics. Circ Res 2003; 92:962-8.
5. Phillips CI, Bogyo M. Proteomics meets microbiology: technical advances in the global mapping of protein expression and function. Cell Microbiol 2005; 7:1061-76.
6. Ahmed N, Rice GE. Strategies for revealing lower abundance proteins in two-dimensional protein maps. J Chromatogr B Analyt Technol Biomed Life Sci 2005; 815:39-50.
7. Fountoulakis M. Application of proteomics technologies in the investigation of the brain. Mass Spectrom Rev 2004; 23:231-58.
8. Cordwell SJ, Basseal DJ, Bjellquist B, and others. Characterization of basic proteins from *Spiroplasma melliferum* using novel immobilized pH gradients. Electrophoresis 1997; 18:1393-8.
9. Corthals GL, Wasinger VC, Hochstrasser DF, JC Sanchez. The dynamic range of protein expression: a challenge for proteomic research. Electrophoresis 2000; 21:1104-15.
10. Chang J, Van Remmen H, Ward WF, and others. Processing of data generated by 2-dimensional gel electrophoresis for statistical analysis: missing data, normalization, and statistics. J Proteome Res 2004; 3:1210-8.
11. Matrix Science. MASCOT. Available from http://www.matrixscience.com/. Accessed 2007 Jul 19.
12. Molecular Station. PepMAPPER. Available from http://www.molecularstation.com/bioinformatics/link/detail/link-453.html. Accessed 2007 Aug 28.
13. ExPASy. Aldente peptide mass fingerprinting. Available from http://ca.expasy.org/tools/aldente/.Accessed 2007 Jul 19.
14. University of California San Francisco. ProteinProspector. Available from http://prospector.ucsf.edu/. Accessed 2007 Jul 19.
15. Mann M, Hendrickson RC, Pandey A. Analysis of proteins and proteomes by mass spectrometry. Annu Rev Biochem 2001; 70:437-73.
16. Domon B, Aebersold R. Mass spectrometry and protein analysis. Science 2006; 312:212-7.
17. Lahm HW, Langen H. Mass spectrometry: a tool for the identification of proteins separated by gels. Electrophoresis 2000; 21:2105-14.

Figure 4. Schematic representation of electrospray ionization-mass spectrometry (ESI-MS) using an ESI-time-of-flight (TOF)-MS instrument



The protein mixture, in liquid phase, is introduced into the MS source through a microsyringe pump, at high voltage. Upon ionization and desolvation, ions are formed. The resulting ions enter the first mass analyzer (MS1), where the molecular masses of individual ion species are determined. In the same mass analyzer, a particular ion can be further selected and transferred to the collision chamber, where it is fragmented by increasing the gas pressure and voltage. The ion fragments are resolved in TOF analyzer and detected as electrons by the photomultiplier detector. The fragmentation spectrum is then manually or automatically processed and interpreted, resulting in the theoretical reconstruction of the original molecule composition and/or structure (Image: Igor C Almeida, University of Texas at El Paso, Department of Biological Sciences).

18. Nyman TA, Rosengren A, Syyrakki S, and others. A proteome database of human primary T helper cells. Electrophoresis 2001; 22:4375-82.
19. Ong SE, Mann M. Mass spectrometry-based proteomics turns quantitative. Nat Chem Biol 2005; 1:252-62.
20. Gygi SP, Rist B, Gerber SA, and others. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol 1999; 17:994-9.
21. Engwegen JY, Gast MC, Schellens JH, Bejinen JH. Clinical proteomics: searching for better tumour markers with SELDI-TOF mass spectrometry. Trends Pharmacol Sci 2006; 27:251-9.
22. Roesch EM, Nees M, Karai S, and others. Transcript and proteome analysis reveals reduced expression of calgranulins in head and neck squamous cell carcinoma. Eur J Cell Biol 2005; 84:431-44.
23. Two-dimensional electrophoresis. Available through Google image search or from http://edoc.hu-berlin.de/dissertationen/xie-jing-2003-12-15/HTML/xie_html_12a07b27.png. Accessed 2007 Jul 19.
24. Lopes, R. Mitochondrial proteome study in canine idiopathic dilated cardiomyopathy and in canine experimental model induced by rapid ventricular pacing. [dissertation]. Urbana (IL) : University of Illinois; 2005.