Introduction to Proteomics

DELFINA C DOMINGUEZ, ROSANA LOPES, M LORRAINE TORRES

Technological advances in the field of genomics have given rise to the development of a new area called proteomics. Proteomics involves the analysis of all proteins expressed in a genome and uses a combination of sophisticated technologies such as two-dimensional electrophoresis, mass spectrometry and bioinformatics to identify and characterize proteins. This new area offers the potential to discover new biomarkers, improve diagnosis, and improve the prognosis of disease processes. This article presents an overview of proteomics importance and related technologies.

ABREVIATIONS: 2-DE = two-dimensional electrophoresis; IPG = immobilized pH gradient; MALDI-TOF = matrix assisted laser desorption ionization time-of-flight; mRNA = messenger ribonucleic acid; pI = isoelectric point.

INDEX TERMS: mass spectrometry, protein expression, proteome, proteomics, two-dimensional electrophoresis.

Clin Lab Sci 2007;20(4):234

***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.*

***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.*

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

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

LEARNING OBJECTIVES

1. Compare and contrast gene expression versus protein expression.
2. Discuss the purpose and goals of proteomics.
3. Identify the steps followed in a proteomic analysis.
4. Describe the main techniques used in proteomics.

The sequencing of the human genome, which unveiled the total genetic content in eukaryotes and advanced technology, has lead to the development of a new field called *proteomics*. Proteomics encompasses the analysis of all proteins expressed in a cell.[1] Two dimensional electrophoresis (2-DE) and mass spectrometry (MS) are the main techniques used in proteomics. Documenting protein expression of an organism by 2-DE analysis is not a new concept. In 1970, Kaltschmidt and Wittmann[2] separated ribosomal proteins from *Escherichia coli* by two-dimensional electrophoresis. A few years later, Klose,[3] O'Farrel,[4] and Scheele[5] described 2-DE as the main technique used for the separation of complex protein mixtures into distinct protein spots. The direct study of protein expression began with Anderson and Anderson in 1977 who for the first time analyzed human plasma proteins and anticipated the possibility of identifying all proteins present in the human body.[6] For the past 20 years, 2-DE has been considered the standard technique for analyzing the proteins expressed by cells, tissues and fluids.[7]

Before the genomics revolution, proteins were sequenced by stepwise chemical degradation from the N terminus to the C terminus followed by UV spectroscopy. These procedures were time consuming and only allowed the identification of few proteins.[8,9] The improvement in efficiency of ionization of molecules by embedding them in a matrix or by electrospray has made possible the rapid identification of thousands of proteins by mass spectrometry.[8,9] The main goals of proteomics are to identify and characterize altered protein expression, investigate co- and post-translational protein modifications, and correlate protein structure and function with biological activity.[10]

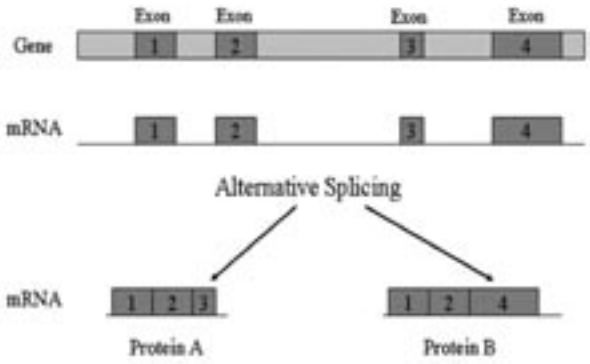
Downloaded from http://hwmaint.clsjournal.ascls.org/ on July 4 2024

WHAT IS PROTEOMICS?

Proteomics is the scientific analysis of all the proteins expressed by a genome.[1] A newly developing subdivision of proteomics is "Clinical Proteomics", which aims to identify proteins involved in pathological processes, to understand how changes in protein expression cause illness, and to develop biomarkers for diagnosis and therapeutic interventions.[10-12] Proteins are implicated in almost all biological functions and for this reason are considered to reflect the true status of a cell. The term proteome was employed for the first time by Wasinger and others (1995) and it refers to the total protein complement of a genome.[1]

There are about 25,000-30,000 genes in the human genome and more than 500,000 predicted proteins in the human proteome.[14,15] These numbers reflect the high degree of complexity of the human genome and clearly show that many genes encode more than one protein. One of the mechanisms that contribute to such complexity is alternative splicing. This process occurs in eukaryotes. Splicing of a transcribed pre-mRNA from one gene can lead to different mature mRNA molecules, and therefore to different proteins. Genes contain both coding DNA (exons) and noncoding DNA (introns). In certain genes, the variable combination of exons causes a gene to produce different proteins.[16,17] An illustration of the alternative splicing process is shown in Figure 1.

Figure 1. Alternative splicing

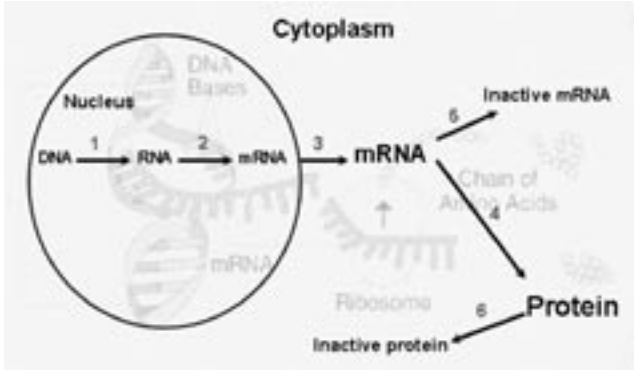


A gene can contain several exons (dark boxes) and introns (light boxes). Exons can be spliced in different ways. For example, if a gene contains 10 exons, one version of the mRNA transcribed from that gene might contain exons 1-9. Another version of the mRNA might contain exons 1-8, and exon 10. Different mature mRNA can lead to different proteins.[31,32]

Eukaryotic organisms have the ability to regulate gene expression by several mechanisms, and at both the transcriptional and post-translational levels. Mechanisms such as RNA splicing, RNA processing, polyadenylation, and the action of regulatory proteins occur at the transcriptional level. These mechanisms control how, where, and when genes are expressed.[17,18] Chemical modifications of proteins like phosphorylation, acetylation, and glycosylation occur at the post-translational level (Figure 2). Post-translational modifications also contribute to the great diversity of proteins and determine cellular localization of proteins, physiological activity, protein interactions, and protein turnover.[19] Therefore, measurements of mRNA expression indirectly correlate with protein function, while protein measurement is related directly to function.[20] These cellular events explain the importance of proteomics. Despite the great degree of complexity that proteomics presents, it offers more specificity than gene analysis alone.

The study of proteomics has been divided into two areas: 1) expression proteomics and 2) functional proteomics. Expression proteomics encompasses the investigation and evaluation of proteins encoded by active genes in a cell, organ, or tissue.[14] Expression proteomics allows the comparison and analysis of differential protein expression profiles between control cells and cells under certain environmental conditions. Therefore, expression proteomics offers the opportunity to compare proteins patterns between health and disease states. On the other hand, functional proteomics analyzes and evaluates protein activity and protein-protein interactions.[14] Func-

Figure 2. Eukaryotic gene expression control points



1) Transcriptional control, 2) RNA processing, 3) RNA transport, 4) translational control, 5) mRNA degradation, 6) Protein activity control[32]

Downloaded from http://hwmaint.clsjournal.ascls.org/ on July 4 2024

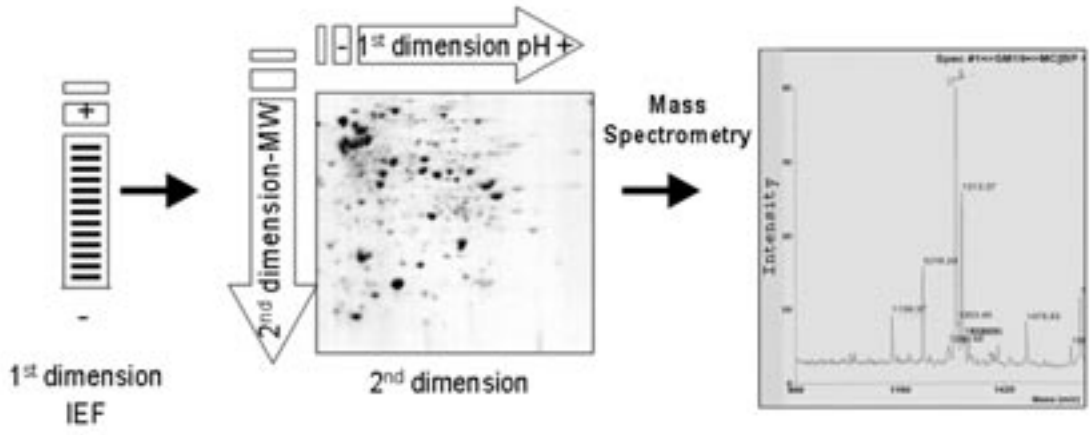
tional proteomics aims to characterize the information flow of intracellular and extracellular protein networks that link tissues and organs.[12] Functional proteomics can be applied to understand signaling mechanisms involving pathological conditions such as cancer, myocardial disease, and brain damage.[21-23] Advances in technology made possible the study of proteomics. Sophisticated instrumentation such as mass spectrometry, improvements in liquid chromatography, and 2-DE apparatus have been essential tools in the identification of proteins.

OVERVIEW OF PROTEOMICS TOOLS

Proteomics uses a combination of several techniques including 2-DE, image analyses, mass spectrometry, amino acid sequencing, and bio-informatics to comprehensively resolve, quantify, and characterize proteins.[7,10,24] Two-dimensional electrophoresis is a method used to separate proteins by their isoelectric point in the first dimension, and subsequently by molecular weight in the second dimension.[7] Proteins resolved in a 2-DE gel can be stained with Coomassie brilliant blue, silver or fluorescent dyes. Proteins spots purified by 2-DE can be excised from the gel and cleaved through chemical or enzymatic methods. Mass spectrometry is the method of excellence for protein identification. The mass differences in the achieved spectrum among the repeated ions are related to the amino acid composition, which can be used to deduce the peptide sequence.[25] The basic steps in a proteomic analysis are illustrated in Figure 3.

2-DE is a versatile technique that can separate hundreds to thousands of proteins at one time (http://ca.expasy.org/ch2d/protocols/).[26] The protein concentration applied to a gel can vary between several micrograms to one milligram. For low abundance proteins to be detected against a background of abundant proteins, such as albumin in a serum sample, a high-protein-capacity system is required. Capacity is dependent on the volume of the gel. Thinner gels provide better sensitivity for the majority of detection methods, and larger and thicker gels offer increased capacity.[27,28] The 2-DE technique is considered an essential method for visualization of proteins according to their molecular weight and isoelectric point. The protein sample should be actively or passively rehydrated before first dimension separation. The first dimension utilizes immobilized pH gradient (IPG) that can be applied to different ranges from broad (pH 3-10) to narrow gradients (pH 3-6, 5-8, and 7-10). Proteins migrate in isoelectric focusing gels to their neutral isoelectric point (pI), where the protein has zero net charge. First-dimensional gels are loaded onto the second-dimensional polyacrylamide gel slabs after equilibration. For the second dimension, polyacrylamide gels are used in different percentages according to molecular weight of proteins to be separated. Several stains can be used to visualize proteins after 2-DE. Stains more commonly used are Coomassie brilliant blue, (detection limit = 8 ng-16ng) and silver staining (detection limit = 2 ng-4ng). In addition, fluorescent dyes can be utilized such as: Sypro Red, Sypro Orange, Sypro Tangerine

Figure 3. Two-dimensional electrophoresis technique



Protein samples are separated according to their isoelectric point (1st dimension) followed by their molecular weight (2nd dimension). Protein identification is commonly performed by mass spectrometry.

Downloaded from http://hwmaint.clsjournal.ascls.org/ on July 4 2024

(detection limit = 4ng-10ng) and Sypro Ruby (detection limit = 1ng-2ng).[29] After staining protein spots of interest, the spots are excised and subjected to enzymatic digestion and subsequently identified by mass spectrometry.

Mass spectrometry (MS) is a powerful technique that identifies unknown compounds and quantifies known compounds and reveals the chemical structure of molecules (http://www.asms.org/whatisms/p1.html).[30] Mass spectrometry measures the mass to charge ratio of ionized molecules and proteins are identified based on their spectrum features.[24] There are different MS techniques used for the identification of proteins. The most commonly used are MALDI-TOF (matrix-assisted laser desorption ionization-time-of-flight), ESI-MS/MS (electrospray ionization tandem mass spectrometry), LC-MS/MS (liquid chromatography tandem mass spectrometry), SELDI (surface enhanced laser desorption ionization), and ICAT (isotope coded affinity tags).[9,14] The principle and applications of these techniques will be discussed in the proteomics technologies section.

CONCLUSION

Advances in genomics (genomic sequences data bases, microarray technology) and technology improvements (mass spectrometry instrumentation) have given rise to a new area of study called proteomics. This field offers the potential to examine and understand protein expression in health and disease, to improve diagnosis, and the possibility to provide a "personalized" therapeutics. This article presents a brief introduction to the field of proteomics presenting history, background information and a review of the technology employed.

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. Colantonio DA, Chan DW. The clinical application of proteomics. Clin Chim Acta 2005; 357:151-8.
2. Kaltschmidt E, Wittmann HG. Ribosomal proteins. XII. Number of proteins in small and large ribosomal subunits of *Escherichia coli* as determined by two-dimensional gel electrophoresis. Proc Natl Acad Sci USA 1970; 67: 1276-82.
3. Klose J. Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. Humangenetik 1975; 26: 231-43.
4. O'Farrell PH. High-resolution two-dimensional electrophoresis of proteins. J. Biol Chem 1975; 250: 4007-21.
5. Scheele GA . Two-dimensional gel analysis of soluble proteins. Characterization of guinea pig exocrine pancreatic proteins. J Biol Chem 1975; 250: 5375-85.
6. Anderson L, Anderson NG. High resolution two-dimensional electrophoresis of human plasma proteins. Proc Natl Acad Sci USA 1977; 74:5421-5.
7. Wilkins MR, Sanchez JC, Gooley AA, and others. Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. Biotechnol Genet Eng Rev 1996;13:19-50.
8. Domon B, Aebersold R. Mass spectrometry and protein analysis. Science 2006;312:212-7.
9. Lahm HW, Langen H. Mass spectrometry: a tool for the identification of proteins separated by gels. Electrophoresis 2000; 21:2105-14.
10. Kavallaris M, Marshall GM. Proteomics and disease: opportunities and challenges. Med J Aust 2005; 182:575-9.
11. Aldred S, Grant MM, Griffiths HR. The use of proteomics for the assessment of clinical samples in research. Clin Biochem 2004; 37:943-52.
12. Petricoin E, Wulfkuhle J, Espina V, Liotta LA. Clinical proteomics: revolutionizing disease detection and patient tailoring therapy. J Proteome Res 2004; 3:209-17.
13. Wasinger VC, Cordwell SJ, Poljak A, and others. Progress with gene product mapping of the mollicutes: *Mycoplasm genitalium*. Electrophoresis 1995; 16: 1090-4.
14. Azad NS, Rasool N, Annunziata CM, and others. Proteomics in clinical trials and practice: present uses and future promise. Mol Cell Proteomics 2006; 5:1819-29.
15. Becker M, Schindler J, Nothwang HG. Neuroproteomics - the tasks lying ahead. Electrophoresis 2006; 27:2819-29.
16. Brett D, Pospisil H, Valcarcel J, and others. Alternative splicing and genome complexity. Nat Genet 2002; 30:29-30.
17. Pennisi E. Why do humans have so few genes? Science 2005; 309:80.
18. Claverie JM. Fewer genes, more noncoding RNA. Science 2005; 309:1529-30.
19. Mann M, Jensen ON. Proteomic analysis of post-translational modifications. Nat Biotechnol 2003; 21:255-61.
20. Anderson NL, Anderson NG. Proteome and proteomics: new technologies, new concepts, and new words. Electrophoresis 1998; 19:1853-61.
21. Borozdenkova S, Westbrook JA, Patel V, and others. Use of proteomics to discover novel markers of cardiac allograft rejection. J Proteome Res 2004; 3:282-8.
22. Fountoulakis M. Application of proteomics technologies in the investigation of the brain. Mass Spectrom Rev 2004; 23:231-58.
23. Vondriska TM, Klein JB, Ping P. Use of functional proteomics to investigate PKC epsilon-mediated cardioprotection: the signaling module hypothesis. Am J Physiol Heart Circ Physiol 2001; 280: H1434-41.
24. Nyman TA, Rosengren A, Syyrakki S, and others. A proteome database of human primary T helper cells. Electrophoresis 2001; 22:4375-82.
25. Gevaert K, Vandekerckhove J. Protein identification methods in proteomics. Electrophoresis 2000; 21:1145-54.

Downloaded from http://hwmaint.clsjournal.ascls.org/ on July 4 2024

26. ExPASy. Technical information on 2-DE gel Available at http://ca.expasy.org/ch2d/protocols/. Accessed 2007 Jul 19.
27. Fichmann J, Westermeier R. 2-D protein gel electrophoresis. An overview. Methods Mol Biol 1999; 112:1-7.
28. Sarmiento M. High-resolution, 2-D protein electrophoresis using nondedicated equipment. Methods Mol Biol 1999; 112:133-45.
29. Berggren K, Chernokalskaya E, Steinberg TH, and others. Background-free, high sensitivity staining of proteins in one- and two-dimensional sodium dodecyl sulfate-polyacrylamide gels using a luminescent ruthenium complex. Electrophoresis 2000; 21:2509-21.
30. American Society for Mass Spectometry. What is mass spectrometry? Available from http://www.asms.org/whatisms/p1.html. Accessed 2007 Aug 28.
31. Alternative splicing. Available from http://www.ncbi.nlm.nih.gov/Class/MLACourse/Modules/MolBioReview/images/alternative_splicing.gif. Accessed 2007 Jul 18.
32. Access Excellence. RNA transcription. Available from http:// www.accessexcellence.org/AE/AEPC/NIH/images/ (rna-transcription.gif) Accessed 2007 Jul 19.

CLINICAL LABORATORY SCIENCE SEEKS MANUSCRIPTS

Clinical Laboratory Science encourages all readers to submit manuscripts for possible publication. We invite submissions to the Dialogue and Discussion, Clinical Practice, and Research and Reports sections.

If you have an idea for an article but aren't sure how to begin, Clinical Laboratory Science editors can put you in touch with a mentor to help you prepare the manuscript. Their contact information, as well as full contact information for the editorial office, can be found on the inside front cover of this issue.

Submission guidelines may be found at http://www.ascls.org/leadership/cls/index.asp. Please contact Margaret LeMay-Lewis, Managing Editor, at ic.ink@mchsi.com with any questions.

Clinical Laboratory Science Announces 2006 Distinguished Author Award Recipients

Recipients of the *Clinical Laboratory Science* Distinguished Author Awards are chosen by *Clinical Laboratory Science* editorial board members. Nominations are based upon based on originality and quality of writing, relevance to the laboratory science profession, and integration of theory and application. The editorial board of *Clinical Laboratory Science* is pleased to announce the following recipients of the 2006 Distinguished Author Awards.

Clinical Practice
Wayne Gade and Brooke Robinson, for their articles *A Brief Survey of Aquaporins and Their Implications for Renal Physiology* and *CLS Meets the Aquaporin Family: Clinical Cases Involving Aquaporin Systems*, published in the Spring 2006 issue of *Clinical Laboratory Science.*

Research and Reports
Rebecca Laudicina, for her article *Searching for Hereditary Hemochromatosis*, published in the Summer 2006 issue of *Clinical Laboratory Science.*

Focus
Emil Scosyrev, for his article *An Overview of the Human Immunodeficiency Virus Featuring Laboratory Testing for Drug Resistance*, published in the Fall 2006 issue of *Clinical Laboratory Science.*

Downloaded from http://hwmaint.clsjournal.ascls.org/ on July 4 2024