

Pediatric proteomics: an introduction

Jeanne Young¹, William L. Stone¹

¹Department of Pediatrics, James H. Quillen College of Medicine, East Tennessee State University, Johnson City, TN 37604

TABLE OF CONTENTS

1. Abstract
2. Introduction
3. Pediatric proteomics and medicine
 - 3.1. Pediatric proteomics is a branch of clinical proteomics
 - 3.2. Proteomics and systems biology
 - 3.3. Clinical proteomics, proteins and biological functions
4. “Omics” and the central dogma of molecular biology
 - 4.1. The original central dogma
 - 4.2. The updated central dogma
 - 4.3. Functional genomics
5. The complexity of human proteins
 - 5.1. What gives rise to an organism’s complexity?
 - 5.2. Protein complexity, gene modularity and post-translational modifications
6. Why Proteomics and Transcriptomics?
 - 6.1 The poor correlation between protein and mRNA abundance
 - 6.2 mRNA sequence cannot predict protein levels or their posttranslational modifications
7. Proteomics provides insight into disease mechanisms
8. Expression proteomics and cell-map proteomics
 - 8.1. The intrinsic complexity of proteomics
 - 8.2. Expression proteomics
9. Advances in proteomic technology
 - 9.1. Gel electrophoresis
 - 9.2. Limitations of 2-D PAGE
 - 9.3. Mass spectrometry (MS), the enabling technology of proteomic
 - 9.4. MALDI, ESI and soft ionization
 - 9.5. MS and Bioinformatics
 - 9.6. MS based proteomics and MudPIT
10. Protein chips and personalized pediatric medicine
 - 10.1. Protein chips
 - 10.2. Future clinical potential
11. Summary
12. Acknowledgements
13. References

1. ABSTRACT

The overall goal of this series is to detail the paradigm shift that proteomics will bring to the practice of pediatric medicine and research. Proteomics is the global study of proteins in a biological system, tissue or bodily fluid. This first review will provide a brief overview of proteomics and describe its niche in the other “omics” of system biology. The underlying technology and methodology will be outlined as well as the obstacles that must be surmounted before pediatric proteomics is optimally useful for clinicians. The potential of proteomics in the area of personalized pediatric medicine will also be discussed since this is of particular clinical relevance. The second article in this series will focus on the application of proteomics to neonatology with particular emphasis on diseases where oxidative stress plays a key pathophysiological role.

2. INTRODUCTION

This review is the first in a series of articles devoted to pediatric proteomics with the goal of providing practitioners with insights into the unique power and clinical usefulness of this branch of systems biology. There are numerous detailed and comprehensive reviews on clinical proteomics but very few specifically devoted to pediatric proteomics (1,2,3). A number of major medical schools are now establishing pediatric proteomics facilities and rapid clinical advances over the next decade are very likely. The first article in this series will provide a brief introduction to the general area of proteomics since this topic is not typically covered in medical school curriculums. The second article in the series will focus on the application of proteomics to neonatology with particular emphasis on diseases where oxidative stress plays a key pathophysiological role. These diseases include

Systems Biology

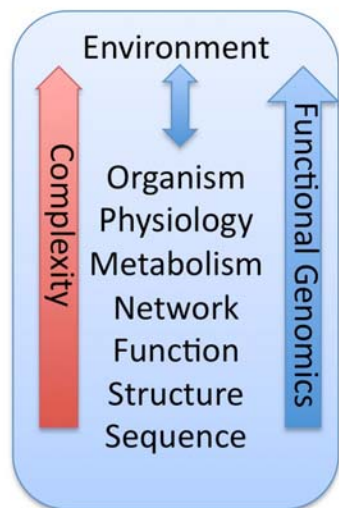


Figure 1. Systems biology utilizes a holistic approach to understand the complex hierarchical organization of an organism and its dynamic interactions with the environment. The underlying technologies that enable this approach are collectively termed “functional genomics.”

retinopathy of prematurity and infant respiratory distress syndrome. Future articles in this series will cover the application of proteomics to pediatric pharmacology and pediatric diseases in general.

3. PEDIATRIC PROTEOMICS AND MEDICINE

The first objective of this review will be to define where “pediatric proteomics” fits into the study and practice of medicine. This will necessitate references to some of the basic concepts of molecular and cellular biology but should not entail additional reading for most readers. Towards this end, most of the references in this review have been selected for their clarity and ability to be read by non-research specialists.

3.1. Pediatric proteomics is a branch of clinical proteomics

Proteomics is concerned with comprehensively identifying and quantifying proteins, determining their functions and interactions with other proteins or macromolecules and characterizing protein-perturbations resulting from development, aging, disease, drug treatment, etc. The term “proteomics” was coined in 1996 and is the protein equivalent of “genomics” or the study of genes (4). While genomics offers powerful insights into disease processes, it is proteins, acting as molecular “nanomachines” that are intimately involved in the molecular causes of disease (5). Clinical proteomics has been elegantly defined by Mischak *et al.* who state that “clinical proteomics is not just a collection of studies dealing with analysis of clinical samples (6). Rather, the essence of clinical proteomics should be to address clinically relevant

questions and to improve the state-of-the-art, both in diagnosis and in therapy of diseases.”

3.2. Proteomics and systems biology

Proteomics is an integral component of systems biology (7). Rather than analyzing the individual components or aspects of an organism's systems, as in traditional molecular biology, systems biology focus on all components and their interactions as part of one system. As indicated in Figure 1, the goal of systems biology is to understand an organism and its environmental interactions as an integrated, dynamic and interacting network of genes, proteins and biochemical reactions. These interactions are ultimately responsible for an organism's form, functions and health. For example, the immune system is not the result of a single mechanism or gene. The immune response arises from the interactions of numerous genes, proteins, and networks as well as the organism's external environment.

The technological platforms that enable a systems biology approach are collectively termed “functional genomics.” Systems biology (see Figure 1) is primarily concerned with integrating information from gene and protein sequences into functional information about proteins and their roles in forming complex signaling and metabolic networks that are dynamically responsive to intra- and extracellular environmental factors. The dynamic aspect of system biology is not easily conveyed by text or figures and covers the molecular motion of protein side chains, on a time scale of 10^{-12} sec, to the life span of an organism, about 2.5×10^9 sec for humans. The reader is strongly encouraged to view the outstanding “The Inner Life of the Cell” videos produced by the Harvard Biovisions groups at Harvard University (see <http://multimedia.mcb.harvard.edu/>) to gain some insight into cellular dynamics.

3.3. Clinical proteomics, proteins and biological functions

As well stated by Zaccai “A protein is a nanomachine whose molecular structure was selected by evolution to perform specific biological functions” (8). It is proteins that primarily perform the functions of living cells, form most subcellular structures (e.g., the cytoskeleton) and act as signaling molecules in information flow. Some key functions of proteins include structural maintenance, enzyme catalysis, immune protection, signal transduction, and the regulation of cell growth and differentiation (9). All of the proteins in a cell constitute the cell's “proteome.” Unlike the relatively static genome, the proteome is constantly changing in response to internal and external signaling events. All disease states perturb the proteome. Studying proteomics therefore gives unique insight into the workings of biological systems and how those systems are influenced by disease and drug therapy. For this reason, clinical proteomics is a powerful complement to genomics in advancing our understanding of human disease, its etiology and treatment. Both drugs and disease states can modulate multiple protein signaling networks and identifying the key proteins in these networks is essential for advancing medical understanding. There is, in fact, a

The Central Dogma and Omics

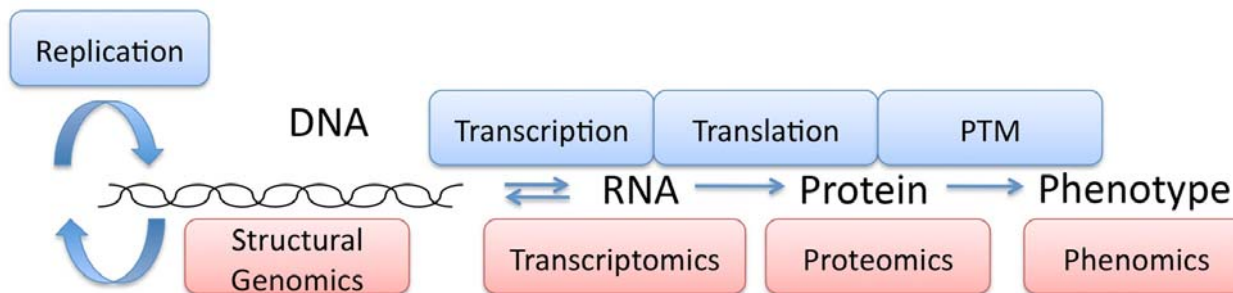


Figure 2. The central dogma of molecular biology forms the foundation for functional genomics, which integrates information from structural genomics, transcriptomics, proteomics and phenomics, i.e., the comprehensive physical and biochemical traits of organisms and the changes caused by genetic mutations and dynamic environmental influences.

relatively new journal called *Current Signal Transduction Therapy* (www.benthamscience.com/cstt/index.htm) solely devoted to identifying and treating signaling disorders. As stated by the editors “In recent years a breakthrough has occurred in our understanding of the molecular pathomechanisms of human diseases whereby most of our diseases are related to intra and intercellular communication disorders.” The breakthroughs have largely been the results of functional genomics.

4.“OMICS” AND THE CENTRAL DOGMA OF MOLECULAR BIOLOGY

4.1. The original central dogma

The sequence of amino acids in a protein is its primary structure and the information dictating this sequence lies within DNA coding sequences. The “central dogma” of molecular biology, as originally described by Crick, holds that the coding information is sequentially and irreversibly transferred from DNA to mRNA (transcription) to protein (translation), i.e., DNA makes RNA makes proteins (10). As indicated in Figure 2, this process begins with the transcription of DNA coding sequences into a complementary strand of messenger ribonucleic acid (mRNA) in the nucleus and terminates in the cytoplasm with the translation of mRNA into the primary protein product (9). Subsequent hydrogen bond formation between the amino acids and hydrophobic interactions lead to an increasingly complicated combination of twists and folds resulting in a polypeptide unit with secondary (primarily alpha-helices and beta-sheets) and tertiary (three-dimensional) structure of the polypeptide units. The polypeptide units can further associate with other identical subunits or with different subunits to form quaternary structures. An outstanding history of protein chemistry can be found in “Nature’s Robots: A History of Proteins (11).

4.2. The updated central dogma

We now know that the original “central dogma” was a simplification over emphasizing the role of structural genomics and the notion that DNA alone was the determining factor controlling cellular functions (12). The informational flow in biological systems is also more

complicated than initially proposed by Crick (10). As originally proposed, the central dogma asserted that information couldn’t be transferred back from proteins to nucleic acids. An example (relevant to clinical issues) where information flows from protein to DNA to mRNA lies in class of proteins called transcription factors. Transcription factors are proteins that bind to specific DNA sequences, modulate the recruitment of RNA polymerase and thereby regulate the expression of adjacent genes. Transcription factors are themselves often regulated by a variety of complex signal transduction pathways that convert external signals into specific cellular responses. As mention above, the “dysregulation” of signal transduction pathways is thought to be critically important in many disease states.

We now also know that retroviruses have a reverse transcription that is able to transfer information from RNA to DNA (the reverse of normal transcription) and hence the left facing arrow in Figure 2. RNA can also make copies of itself. Systems biology does, however, heavily rely on sequence data to determine how genes work in an organism dynamically interacting with its environment.

4.3. Functional genomics

The technologies employed by systems biology are shown in Figure 2 (structural genomics, transcriptomics, proteomics, etc.) and they are collectively called functional genomics. Technologies such as structural genomics, transcriptomics and proteomics provide “part lists” for the macromolecular components of a biological system and it is the job of systems biology to address the issue of how these parts fit together into functional and dynamic networks that are responsive to environmental stimuli. Excellent software providing a graphical notation system for representing models of biochemical and gene-regulatory networks is now freely available (13).

5. THE COMPLEXITY OF HUMAN PROTEINS

5.1. What gives rise to an organism’s complexity?

The human genome was sequenced in its entirety by 2001 and it was anticipated that this major

accomplishment would lay the foundations “for ongoing and future endeavors that will revolutionize biomedical research and our understanding of human health” (14,15). Nevertheless, only about 21,000 protein-encoding transcripts were found in the human genome: a surprising discovery considering it represents a mere one third increase over the number of genes in the nematode, one of the most basic multicellular organisms in existence (16). The question arises, how can an organism as complex as a human function with a gene complement not much larger than that of a microscopic worm? The answer may lie in the complexity of human proteins. Compared to about 21,000 genes, the human body has been estimated to have over 200,000 proteins with some estimates closer to one million.

5.2. Protein complexity, gene modularity and post-translational modifications

Although controversial, it is possible that human “complexity” could arise from a greater modularity of genes compared with that of other multicellular organisms (17,18). Alternative splicing of messenger RNA (mRNA) permits many gene products (i.e., proteins) to be made from a single coding sequence. More specifically, alternative splicing is the process by which introns (non-coding DNA) are removed from an RNA precursor (pre-mRNA) and remaining exons are linked together to form a mature mRNA. Pre-mRNA has the potential to be processed into multiple mRNAs with each mRNA being translated into a unique protein polypeptide. Genes with greater “modularity” would provide a more complicated set of proteins. With alternative mRNA splicing, 21,000 genes can easily encode four times as many proteins.

Chemical modification of proteins, after their initial synthesis (post-translational modifications or PTMs), also increases complexity (see Figure 2) by modifying the functions of proteins and their interactions with other macromolecules. Issues surrounding the complexity of all the RNA transcripts produced by an organism, i.e., the transcriptome, remains an active area of research (18). A full understanding of an organism’s complexity is dependent on a comprehensive understanding of proteins, their interaction with each other and other biomolecules (i.e., the interactome) as well as an ever changing set of environmental signals. This is the domain of the relatively new discipline of proteomics.

6. WHY PROTEOMICS AND TRANSCRIPTOMICS?

6.1. The poor correlation between protein and mRNA abundance

Since mRNA represents the first step in gene expression it is reasonable to suggest that studying gene transcripts (mRNA) alone could be sufficient to provide a complete picture of gene expression in cells. The availability of modern high-throughput DNA microarray technology permits the simultaneous quantitative analysis of mRNA expression for thousands of genes, which is a compelling advantage of transcriptomics. A key issue is how well the levels of an mRNA transcript reflect the levels of its translated protein.

Guo *et al.* recently addressed the question of whether or not mRNA expression is good predictor of protein expression in humans (19). These researchers found a significant but weak ($r=0.235$) correlation between mRNA and protein expression for the 71 genes examined. As discussed by Greenbaum *et al.* the measurement of mRNA levels have proven to be clinically useful since they often correlate with disease states (20). Greenbaum *et al.* state, however, that “these results are almost certainly correlative, rather than causative: in the end it is most probably proteins and their interactions that are the true causative forces in the cell, and it is the corresponding protein quantities that we ought to be studying” (20).

There are several proposed reasons for the observed lack of robust correlation between an mRNA and its translated protein. Fundamentally, the molecular regulatory mechanisms governing the half-lives of mRNA and protein are different. Some mRNA’s are transcribed but not translated into protein or are not efficiently translated. This follows from the finding that mRNA molecules are relatively unstable with their primary function being the transmission of a genetic message from the nucleus of the cell to the protein making machinery of the cytoplasm. As the target of RNases, mRNAs are rapidly degraded to prevent the overproduction of proteins. This degradation, as a regulatory control mechanism of the cell, can occur with or without the concurrent translation of mRNA to protein thereby contributing to the lack of correlation between an mRNA and its protein product.

6.2. mRNA sequence cannot predict protein levels or their posttranslational modifications

The post-transcriptional events and dynamics that ultimately turn an mRNA into a protein are not sufficiently characterized to even permit a quantitative estimate of the amount of protein produced. Moreover, post-translational modifications to proteins themselves, such as phosphorylation and glycosylation, can change the secondary, tertiary or quaternary structure of a protein and thereby alter its function(s). For this reason, the mRNA sequence may not be representative of the final, mature protein, i.e., its phenotype. Neither structural genomics nor transcriptomics hold definitive information from which to deduce subsequent protein modifications. The importance and complexity of protein PTMs have led to the development of proteomic subspecialty fields such as phosphoproteomics and glycoproteomics. The control of gene expression in the cell can also occur at the protein level. Proteins are degraded at varying rates by cellular enzymes in response to various cellular signals. These changes in protein activity are not predictable from the DNA/RNA sequences.

7. PROTEOMICS PROVIDES INSIGHT INTO DISEASE MECHANISMS

Disease represents perturbations in the normal functioning of biological systems. Because proteins are the embodiment of cellular activity, analysis of protein structure, function, interactions and expression can provide unique insights into disease mechanisms that pure gene-

The Two Branches of Proteomics

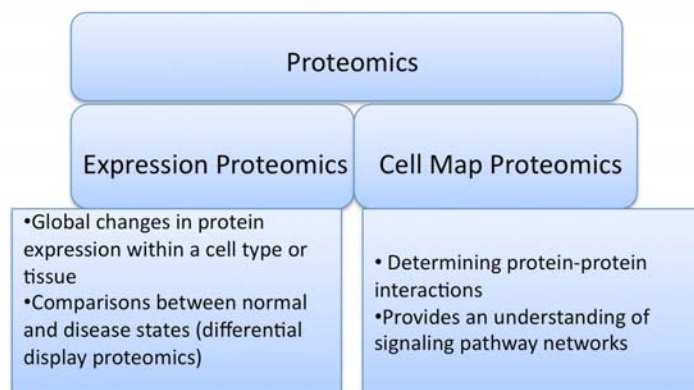


Figure 3. Proteomics has two primary branches, i.e., expression proteomics and cell map proteomics.

based research cannot offer. Most drug targets are proteins and the interacting signaling networks formed by proteins: proteomics therefore provides the potential for discovering novel pharmaceutical targets and individualizing drug therapy.

The dynamic activity of a biological system is paralleled by dynamic changes in protein expression. As a cell responds to various internal and external stimuli, the proteins continually change in order to meet the varying needs of the cell and to maintain homeostasis. By examining the proteins present in a cell, tissue or fluid sample, as well as the changes in that protein expression, a more comprehensive picture of cellular activity can be achieved as well as an increased understanding of the mechanisms by which these activities are taking place. Disease states are almost universally associated with changes in protein expression: whether these changes are the cause or the consequence of the disease state they can nevertheless can provide important clinical information for diagnoses and management.

8. EXPRESSION PROTEOMICS AND CELL-MAP PROTEOMICS

8.1. The intrinsic complexity of proteomics

A major problem of proteomics lies in its intrinsic complexity: it has been estimated that amount of data derived from characterizing all human proteins is at least three-orders of magnitude greater than that of the human genome project. Acquiring, analyzing and interpreting this vast amount of information remains a daunting challenge. Moreover, the development of high throughput technologies in proteomics is not as advanced as in genomics: proteomics also lacks an “amplification” technique as afforded by the polymerase chain reaction (PCR) for DNA.

In order to structure the complex information derived from applied proteomics it is useful to consider (as shown in Figure 3) two major proteomic subdivisions, i.e., expression proteomics and cell-map proteomics (21).

8.2. Expression proteomics

In expression proteomics the global level of protein expression is measured in a biological sample (e.g., serum, bronchoalveolar lavage fluid). Most often, the differences in protein expression (differential expression proteomics) between clinically relevant situations are measured. This approach is particularly useful for comparing normal and disease states and thereby identifying potential candidate biomarkers for the disease state that could be used for early detection or drug targets. Two-dimensional gel electrophoresis (see below) followed by protein identification using mass spectrometry (MS) is the mainstay of expression proteomics.

8.3. Cell map proteomics

The second subdivision, cell map proteomics, provides a more interactive picture of protein expression and involves identifying protein location and protein-protein interactions/networks within cells. Because many proteins act as parts of multi-protein cellular complexes, identification of protein location and interactions allows for the formation of more complex functional protein maps, the ‘cell-map’. This information is useful in deciphering protein function and in identifying new drug targets.

9. ADVANCES IN PROTEOMIC TECHNOLOGY

As mentioned above, the main challenge in proteomic research has been the lack of efficient, high throughput methodology with which to investigate protein expression in biological systems. Due to the high complexity of protein expression in a sample as well as its constantly changing nature, instrumentation with which to image and identify large numbers of proteins quickly is critical. Ongoing advances in mass spectrometry, such as those implemented by John Fenn and Koichi Tanaka, who received the 2002 Nobel Prize for Chemistry, have gone far to making this a reality. Prior to their work, only relatively small molecules could be identified by mass spectrometry but they developed the technology for analyzing macromolecules like proteins (22).

9.1. Gel electrophoresis

The first stage in proteomic workflow usually involves the separation of proteins from a biological sample such as serum, urine or bronchoalveolar (BAL) fluid. Two-dimensional polyacrylamide gel electrophoresis (2-D PAGE) was the first method used for proteome analysis and remains a widely used and effective technique for separating large protein mixtures. At present, there is no other technology capable of resolving thousands of proteins in one separation procedure. In this method, proteins are denatured with urea (neutral charge) and a non-ionic or zwitter-ionic detergent, and the polypeptides are first separated (1st-dimension) according to their isoelectric points and then (2nd-dimensions) by their “apparent” molecular weights. These two sequential separation steps occur at 90 degree from each other in the gel and hence the term “2-dimensional”. Highly sensitive protein staining methods allow visualization of the separated polypeptides on the 2-D PAGE gels. The pattern of staining on the gels can then be compared between samples or against standard sample maps. Databases of detailed gel maps of body fluids such as human plasma or BAL are currently available. Proteins of interest can be removed from the gel and further characterized using mass spectrometry.

Even small experimental variations between individual gels can cause problems when trying to compare 2-D PAGE gels between “normal” and “disease” states. This issue has now been largely overcome through the use of two-dimensional difference gel electrophoresis (2D-DIGE). Using sample-specific fluorescent labeling, 2D-DIGE allows the simultaneous superimposition of multiple protein extracts on the same gel thereby removing the problem of experimental variation between individual gels run separately. This technique relies on labeling each protein extract with a spectrally resolvable fluorescent dye: scanning the gel with different wavelengths produces distinct images for each protein extract (23). An internal standard, which is a mixture of all the samples used in a given experiment, is essential and permits DIGE to quantify the amount of each resolved polypeptide

9.2. Limitations of 2-D PAGE

Despite being the “core technology” of expression proteomics, 2-D PAGE has its limitations: (1) it generally does not work with membrane bound proteins which form lipid-protein complexes that are not easily solubilized in the 1st-dimensional run; (2) is of limited value for low abundant proteins; (3) analysis and quantification can be difficult; (4) native protein structure and function are lost due to the requirement for protein denaturation. While 2-D PAGE is excellent for the quantitative analysis of complex protein samples there is, nevertheless, a tendency to under-represent certain classes of proteins. Classically, high or low molecular weight proteins, membrane proteins and proteins with extreme isoelectric points have been difficult to isolate with 2-D PAGE. The differentially expressed spots of the gel must be further analyzed by mass spectrometry and manually excising these spots is slow and labor intensive. Automating this process requires a very expensive robotic “spot-picker.” High abundance proteins must also be

removed since they mask the presence of proteins present in smaller quantities. Moreover, DIGE is not as sensitive as pure MS-based approaches (as discussed below).

9.3. Mass spectrometry (MS), the enabling technology of proteomics

MS is an analytical technique that precisely measures the mass (m)-to-charge (z) ratio (m/z) for molecules in the gas phase. After proteins are separated by 2-D-PAGE or liquid chromatography (LC), the next important proteomic goal is protein identification, which is almost always achieved by MS. Current technology requires that the protein in a sample first be digested with a protease to break the protein into peptide fragments. Proteases split the peptide bond linking amino acid residues in the polypeptide chain. Using an amino acid specific protease, such as trypsin, is useful since it enables the specific peptides formed proteolysis to be theoretically predictable. For 2-D-PAGE, the protein spots are subjected to in-gel protein digestion, the peptide sections are ionized, vaporized and introduced into the MS, which detects their m/z ratio. Since peptides are thermally labile and not particularly volatile this has been a challenging task.

9.4. MALDI, ESI and soft ionization

Two techniques have, however, made MS based proteomics possible since they utilize “soft” ionization that transfer the intact peptide (or protein) into the gaseous phase without molecular disruption. The first is matrix-assisted laser desorption/ionization (MALDI), which uses a laser to vaporize and ionize the molecules in a sample dried on a metal plate within a chemical matrix. The chemical matrix is used to protect the molecules from being destroyed by the laser and to facilitate vaporization and ionization. Electrospray ionization (ESI) is the second “soft” ionization method and it forces the sample to flow through a small, charged capillary tube, ionizing the sample and spraying it out in an aerosol. Both MALDI and ESI can detect low protein levels and are suitable for automation. Compared with MALDI, ESI is a lower throughput technique. Moreover, with ESI the sample going into the MS is typically from a high performance liquid chromatography (HPLC) system and is consumed, i.e., the sample cannot be reanalyzed as can be done with MALDI. However, ESI has the advantage of being able to study larger peptides/proteins since it typically produces multiply charged ions with a m/z ratio that fall in the highly resolvable range of most MS. ESI is also the preferred method for the studying protein post-translational modifications.

9.5. MS and bioinformatics

The next key step in proteomic workflow is using the MS data for protein identification (Figure 4). For MALDI based MS, a peptide mass fingerprinting (PMF) technique is used where the m/z ratios for peptides in the unknown protein (the “fingerprint”) are compared to a database containing all the known protein sequences in the genome of the species being investigated. All the proteins in the database are “theoretically” cleaved with the protease used for sample digestion into peptides with calculable m/z ratios that are statistically compared with the observed m/z

MS-Based Proteomics

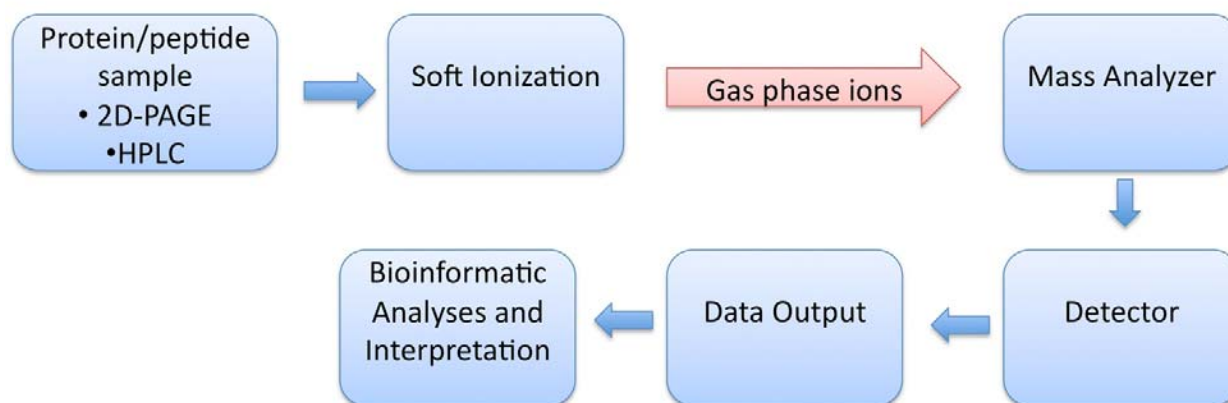


Figure 4. A mass spectrometer (MS) measures the mass to charge ratio (m/z) of an ionized molecule. For MS based-proteomic analyses the protein sample is subject to a soft ionization technique (MALDI or ESI), which produces gas phase protein/peptide ions that sequentially enter a mass analyzer and a mass detector. The MS output data is further analyzed and interpreted by bioinformatic software.

ratios from the unknown protein to find the best protein match.

9.6. MS based proteomics and MudPIT

MS-based proteomics combined with HPLC protein separation technology is an attractive alternative to gel-based proteomics (such as 2-DIGE) due to its increased sensitivity and ability to identify proteins with low abundance, high or low molecular weight, extreme hydrophobicity or isoelectric point. “Shotgun proteomics”, named after shotgun DNA sequencing, is the key enabling methodology for identifying proteins in MS/HPLC-based approaches. In this method, a complex protein mixture is first digested with a sequence-specific protease (typically trypsin) yielding an even more complex mixture of peptides that are then analyzed by MS (see ref (24) for a very lucid and detailed description). Each peptide is isolated in the MS and subjected to tandem MS (MS/MS) that provides the actual sequence of amino acids in the peptide. These data are then used to reconstruct the identity of parent proteins from which the complex mixture of peptides is derived. Needless to say, this is a challenging task and a powerful testimony to the skill of bioinformatic programmers and the power of modern computers. The peptides produced by proteolytic digestion are separated by HPLC before being introduced into the MS by electrospray ionization. In order to obtain even better separation, two-dimensional LC can be employed with the first LC separation being performed on a strong cationic column and the second separation on a reverse phase column. This LC/LC approach is termed multidimensional protein identification technology (MudPIT) and it can separate and identify many thousands of peptides in a sample by the sequential application of electrospray ionization, MS/MS, and database searching (25).

Recent advances have improved the ability of MudPIT to provided quantitative information. In 2D-DIGE,

protein quantification is done at the whole protein level and information about protein isoforms with posttranslational modification is retained (very useful). With MudPIT, quantification is done at the level of peptides rather than at the protein level as in 2D-DIGE: this makes it difficult to obtain quantitative information on posttranslational modifications from MudPIT. Some proteomic centers perform both MudPIT and 2D-DIGE since these approaches yield complementary information.

10. PROTEIN CHIPS AND PERSONALIZED PEDIATRIC MEDICINE

10.1. Protein chips

The technology detailed above is usually not available in a clinical setting and is both expensive and labor intensive. Mallick and Kuster (26) recently published an excellent review of proteomics, from a pragmatic perspective, emphasizing the notion that there is no “one size fits all” proteomic strategy for all biological questions.

For pediatric proteomics there is a need for high sensitivity since only very small samples of biofluids are available for analysis. Moreover, turn around time for sample analysis must be very short for the data to be optimally useful for immediate clinical guidance. Ideally, the sample turn around time should be minutes rather than the days typical of a proteomic core facility. In addition to rapid turn around time, there must be a comprehensive and rapid clinical analysis of the proteomic results in a language useful to clinical providers.

10.2. Future clinical potential

Protein chips represent a technology that has enormous practical clinical potential and overcomes many of the issues raised above. Protein chips, which are also called protein microarrays, hold the promise of performing

large-scale high throughput proteomic analyses of biofluids using cost-effective technology with clinical personal rather than research-trained specialists.

Typically, a protein chip consists of “capture” molecules immobilized in rows and columns on a flat surface (27). The capture molecules typically bind fluorescently labeled target proteins in the biofluid sample and the captured target proteins are then visualized by fluorescence. The power of protein chip technology lies in its ability to measure a large number of analytes simultaneously.

The ideal protein chip could be used at the “bedside” by a primary care physician, physician’s assistant or nurse practitioner with a biofluid such as BAL, plasma or urine. The dipstick test commonly used for urinalysis could be considered a “primitive” precursor to a protein chip. The dipstick is a very cost-effective tool that can detect abnormalities in the urinary system as well as other organ systems, including liver function, acid-base status, and carbohydrate metabolism. The protein chip equivalent of the dipstick could represent a quantum leap in terms of number of biomolecules being analyzed and the diagnostic information that could be generated. If appropriately designed, the signals from the protein microarray could be digitized, sent to the Internet via a smartphone, the results analyzed, interpreted and sent back to the health care provider in a matter of minutes.

Although the “ideal” protein chip has not yet been commercialized, a prototype chip described by Evans *et al.* comes close to the mark (28). These investigators used immobilized protein aptamers rather than the usual immobilized antibodies as “capture” molecules. Peptide aptamers have a variable peptide domain that is “tethered” to a stable scaffold protein that reduces the set of confirmations available to the variable domain of the peptide aptamers. The variable domain binds to a specific non-denatured target protein. Peptide aptamers are similar to antibodies where the heavy and light chains form a “scaffold” but are considerably less fragile and have more specific binding to the target proteins after being immobilized on a surface. Evans *et al.* have also devised an electronic “label-free” strategy that does away with the necessity of fluorescently labeling the proteins in the biofluid sample (28). Finally, these investigators have employed conventional silicon micro-fabrication technology to produce a high-density microarray chip with “integrated readout technology capable of performing the many simultaneous measurements required for proteome-wide studies” (28).

The biomolecules to be measured by a protein chip would be limited to those that are validated biomarkers for disease prognosis, diagnosis or useful in predicting which individuals will respond to a given therapeutic regimen, i.e., personalized medicine. 2-D PAGE and MudPIT remain, however, the technologies required to initially identify the candidate biomarkers.

11. SUMMARY

This first article has detailed the unique niche that pediatric proteomics occupies in the larger scheme of

functional genomics, clinical proteomics and systems biology. In addition, a brief introduction into the underlying technology of proteomics has been provided along with a description of the limitations and future possibilities. The next article in this series will focus on the application of pediatric proteomics to the area of neonatology with emphasis diseases in which oxidative stress plays a key role.

12. ACKNOWLEDGEMENTS

A preliminary draft of this review was written by Jeanne Young, a medical student at the James H Quillen College of Medicine and the final version subsequently completed by William L. Stone, Director of Pediatric Research at the James H Quillen College of Medicine. Ms. Young received financial support from the Summer Research Fellowship Program at the James H Quillen College of Medicine.

13. REFERENCES

1. P. Lescuyer, A. Farina and D. F. Hochstrasser: Proteomics in clinical chemistry: will it be long? *Trends Biotechnol* 28, 225-229 (2010)
2. E. F. Petricoin and L. A. Liotta: Clinical proteomics: application at the bedside, *Contrib Nephrol* 141, 93-103 (2004)
3. S. W. Hunsucker, F. J. Accurso and M. W. Duncan: Proteomics in pediatric research and practice, *Adv Pediatr* 54, 9-28 (2007)
4. N. L. Anderson and N. G. Anderson: Proteome and proteomics: new technologies, new concepts, and new words, *Electrophoresis* 19, 1853-1861 (1998)
5. T. Nishimura, A. Ogiwara, K. Fujii, T. Kawakami, T. Kawamura, H. Anyouji and H. Kato: Disease proteomics toward bedside reality, *J Gastroenterol* 40 Suppl 16, 7-13 (2005)
6. H. Mischak, R. Apweiler, R. E. Banks, M. Conaway, J. Coon, A. Dominiczak, J. H. Ehrich, D. Fliser, M. Girolami, H. Hermjakob, D. Hochstrasser, J. Jankowski, B. A. Julian, W. Kolch, Z. A. Massy, C. Neusuess, J. Novak, K. Peter, K. Rossing, J. Schanstra, O. J. Semmes, D. Theodorescu, V. Thongboonkerd, E. M. Weissinger, J. E. Van Eyk and T. Yamamoto: Clinical proteomics: A need to define the field and to begin to set adequate standards, *Proteomics Clin Appl* 1, 148-156 (2007)
7. T. Ideker, Systems Biology 101-what you need to know, *Nature Biotechnology* 22, 473 (2004)
8. G. Zaccai, Proteins and nano-machines: dynamics-function relations studied by neutron scattering, *J Phys :Condens Matter* 15, S1673-S1682 (2003)
9. J. D. Pierce, M. Fakhari, K. V. Works, J. T. Pierce and R. L. Clancy: Understanding proteomics, *Nurs Health Sci* 9, 54-60 (2007)

10. F. Crick, Central dogma of molecular biology, *Nature* 227, 561-563 (1970)
11. C. Tanford and J. Reynolds: Nature's Robots: A History of Proteins. USA: *Oxford University Press*, (2004)
12. J. A. Shapiro, Revisiting the central dogma in the 21st century, *Ann N Y Acad Sci* 1178, 6-28 (2009)
13. H. Kitano, A. Funahashi, Y. Matsuoka and K. Oda: Using process diagrams for the graphical representation of biological networks, *Nat Biotechnol* 23, 961-966 (2005)
14. G. Subramanian, M. D. Adams, J. C. Venter and S. Broder: Implications of the human genome for understanding human biology and medicine, *JAMA* 286, 2296-2307 (2001)
15. J. C. Venter, M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, J. D. Gocayne, P. Amanatides, R. M. Ballew, D. H. Huse, J. R. Wortman, Q. Zhang, C. D. Kodira, X. H. Zheng, L. Chen, M. Skupski, G. Subramanian, P. D. Thomas, J. Zhang, G. L. Gabor Miklos, C. Nelson, S. Broder, A. G. Clark, J. Nadeau, V. A. McKusick, N. Zinder, A. J. Levine, R. J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. Di Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A. E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T. J. Heiman, M. E. Higgins, R. R. Ji, Z. Ke, K. A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G. V. Merkulov, N. Milshina, H. M. Moore, A. K. Naik, V. A. Narayan, B. Neelam, D. Nusskern, D. B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Wang, A. Wang, X. Wang, J. Wang, M. Wei, R. Wides, C. Xiao, C. Yan, A. Yao, J. Ye, M. Zhan, W. Zhang, H. Zhang, Q. Zhao, L. Zheng, F. Zhong, W. Zhong, S. Zhu, S. Zhao, D. Gilbert, S. Baumhueter, G. Spier, C. Carter, A. Cravchik, T. Woodage, F. Ali, H. An, A. Awe, D. Baldwin, H. Baden, M. Barnstead, I. Barrow, K. Beeson, D. Busam, A. Carver, A. Center, M. L. Cheng, L. Curry, S. Danaher, L. Davenport, R. Desilets, S. Dietz, K. Dodson, L. Doup, S. Ferreira, N. Garg, A. Gluecksmann, B. Hart, J. Haynes, C. Haynes, C. Heiner, S. Hladun, D. Hostin, J. Houck, T. Howland, C. Ibegwam, J. Johnson, F. Kalush, L. Kline, S. Koduru, A. Love, F. Mann, D. May, S. McCawley, T. McIntosh, I. McMullen, M. Moy, L. Moy, B. Murphy, K. Nelson, C. Pfannkoch, E. Pratts, V. Puri, H. Qureshi, M. Reardon, R. Rodriguez, Y. H. Rogers, D. Romblad, B. Ruhfel, R. Scott, C. Sitter, M. Smallwood, E. Stewart, R. Strong, E. Suh, R. Thomas, N. N. Tint, S. Tse, C. Vech, G. Wang, J. Wetter, S. Williams, M. Williams, S. Windsor, E. Winn-Deen, K. Wolfe, J. Zaveri, K. Zaveri, J. F. Abril, R. Guigo, M. J. Campbell, K. V. Sjolander, B. Karlak, A. Kejariwal, H. Mi, B. Lazareva, T. Hatton, A. Narechania, K. Diemer, A. Muruganujan, N. Guo, S. Sato, V. Bafna, S. Istrail, R. Lippert, R. Schwartz, B. Walenz, S. Yoosheph, D. Allen, A. Basu, J. Baxendale, L. Blick, M. Caminha, J. Carnes-Stine, P. Caulk, Y. H. Chiang, M. Coyne, C. Dahlke, A. Mays, M. Dombroski, M. Donnelly, D. Ely, S. Esparham, C. Fosler, H. Gire, S. Glanowski, K. Glasser, A. Glodek, M. Gorokhov, K. Graham, B. Gropman, M. Harris, J. Heil, S. Henderson, J. Hoover, D. Jennings, C. Jordan, J. Jordan, J. Kasha, L. Kagan, C. Kraft, A. Levitsky, M. Lewis, X. Liu, J. Lopez, D. Ma, W. Majoros, J. McDaniel, S. Murphy, M. Newman, T. Nguyen, N. Nguyen, M. Nodell, S. Pan, J. Peck, M. Peterson, W. Rowe, R. Sanders, J. Scott, M. Simpson, T. Smith, A. Sprague, T. Stockwell, R. Turner, E. Venter, M. Wang, M. Wen, D. Wu, M. Wu, A. Xia, A. Zandieh and X. Zhu: The sequence of the human genome, *Science* 291, 1304-1351 (2001)
16. J. M. Claverie, Gene number. What if there are only 30,000 human genes? *Science* 291, 1255-1257 (2001)
17. D. Brett, H. Pospisil, J. Valcarcel, J. Reich and P. Bork: Alternative splicing and genome complexity, *Nat Genet* 30, 29-30 (2002)
18. S. Gustincich, A. Sandelin, C. Plessy, S. Katayama, R. Simone, D. Lazarevic, Y. Hayashizaki and P. Carninci: The complexity of the mammalian transcriptome, *J Physiol* 575, 321-332 (2006)
19. Y. Guo, P. Xiao, S. Lei, F. Deng, G. G. Xiao, Y. Liu, X. Chen, L. Li, S. Wu, Y. Chen, H. Jiang, L. Tan, J. Xie, X. Zhu, S. Liang and H. Deng: How is mRNA expression predictive for protein expression? A correlation study on human circulating monocytes, *Acta Biochim Biophys Sin (Shanghai)* 40, 426-436 (2008)
20. D. Greenbaum, C. Colangelo, K. Williams and M. Gerstein: Comparing protein abundance and mRNA expression levels on a genomic scale, *Genome Biol* 4, 117 (2003)
21. W. P. Blackstock and M. P. Weir: Proteomics: quantitative and physical mapping of cellular proteins, *Trends Biotechnol* 17, 121-127 (1999)
22. J. Koomen, D. Hawke and R. Kobayashi: Developing an understanding of proteomics: an introduction to biological mass spectrometry, *Cancer Invest* 23, 47-59 (2005)
23. R. Marouga, S. David and E. Hawkins: The development of the DIGE system: 2D fluorescence difference gel analysis technology, *Anal Bioanal Chem* 382, 669-678 (2005)
24. E. M. Marcotte, How do shotgun proteomics algorithms identify proteins? *Nat Biotechnol* 25, 755-757 (2007)
25. C. M. Delahunty and J. R. Yates 3rd: MudPIT: multidimensional protein identification technology, *BioTechniques* 43, 563, 565, 567 passim (2007)
26. P. Mallick and B. Kuster: Proteomics: a pragmatic perspective, *Nat Biotechnol* 28, 695-709 (2010)

27. X. Yu, N. Schneiderhan-Marra and T. O. Joos: Protein microarrays for personalized medicine, *Clin Chem* 56, 376-387 (2010)

28. D. Evans, S. Johnson, S. Laurenson, A. G. Davies, P. Ko Ferrigno and C. Walti: Electrical protein detection in cell lysates using high-density peptide-aptamer microarrays, *J Biol* 7, 3 (2008)

Abbreviations: PMT: posttranslational modification; RNase: ribonuclease; PCR: polymerase chain reaction; MS: mass spectrometry; BAL: bronchoalveolar lavage; 2-D-PAGE: two-dimensional polyacrylamide gel electrophoresis; 2D-DIGE: two-dimensional difference gel electrophoresis; LC: liquid chromatography; MALDI: matrix-assisted laser desorption/ionization; ESI: electrospray ionization; HPLC: high performance liquid chromatography; MudPIT: multidimensional protein identification Technology

Key words: Pediatrics, Proteomics, Systems Biology, Protein Chip, Medicine, Review

Send correspondence to: William L. Stone, James H. Quillen College of Medicine, East Tennessee State University, Johnson City, TN 37614 USA, Tel: 423-439-8762, Fax: 423-439-8066, E-mail: stone@etsu.edu

<http://www.bioscience.org/current/vol4S.htm>