Potential Applications and Limitations of Proteomics in the Study of Neurological Disease

Yoshito Kinoshita, PhD; Takuma Uo, PhD; Suman Jayadev, MD; Gwenn A. Garden, MD, PhD; Thomas P. Conrads, PhD; Timothy D. Veenstra, PhD; Richard S. Morrison, PhD

Proteomics represents the comprehensive study of cellular proteins and is aimed at analyzing their structure, function, expression, interactions, and localization in complex biological systems. The information obtained from these types of analyses can contribute to our understanding of the function of individual proteins by identifying protein-protein interactions and dynamic protein networks found in normal and diseased conditions. Genomic (DNA) or transcriptomic (messenger RNA) approaches alone do not take into account changes in protein stability, localization, and posttranslational modifications that are often critical determinants of protein function and, by extension, cellular behavior. Although proteomic methods still require significant technical advances to provide a truly “global” or “comprehensive” measure of gene expression similar to that achieved by DNA microarrays, recent advances in proteomics are beginning to provide a means to simultaneously characterize the expression of thousands of proteins in a whole cell or bio-fluid proteome and hundreds of proteins in select subcellular structures or protein complexes. The information obtained from these studies should promote a better understanding of disease conditions, help therapeutic decision making, and potentially foster the identification of therapeutic targets by comparing the proteomes of normal and diseased samples.

Two major methods are available to identify proteins on a proteomewide scale using mass spectrometry (MS). The classical method using 2-dimensional polyacrylamide gel electrophoresis (2-DE) separates complex mixtures of proteins based on their molecular mass and isoelectric point. Enzyme digestion of the resolved protein spots is then used so that the masses of the resulting peptides can be measured using MS for protein identification (2-DE–MS). Although 2-DE–MS is an established method, many proteins are incompatible with separation by 2-DE or cannot be detected on a gel with sufficient sensitivity. This in-gel separation is also difficult to integrate online with MS analysis, thus limiting its use for high-protein coverage and high-throughput proteomics. Alternatively, a “shotgun” proteomics approach has been developed in which an enzyme-digested protein sample is separated in solution using liquid chromatography (LC) and is directly coupled online to sequence-based protein identification by tandem mass spectrometry (MS/MS). In MS/MS, an ionized peptide peak is first selected from the mixture of ions produced in the source region of the mass spectrometer. The selected molecular ion is then further fragmented into smaller ions. Measuring the masses of the fragment ions allows identification of the peptide based on partial amino acid sequence information. Peptides are identified by using various methods to match the MS/MS spectrum to a database of protein sequences. This LC-MS/MS approach facilitates higher proteome coverage and more sensitive and higher throughput analysis compared with the 2-DE–MS method.
method. Presently, further physicochemical prefractionation of a protein sample is required to reduce its complexity prior to LC-MS/MS and thereby achieve identification of low-abundance proteins. Using this latter approach, 4000 to 5000 proteins can currently be identified in a whole cell or serum proteome.

Proteome coverage can also be dramatically improved by analyzing a discrete subset of a proteome (subproteome), which reduces the complexity of a biological sample prior to MS analysis. This strategy has been applied with great success to a variety of subcellular structures including mitochondria, and plasma membrane. Approaches are also being developed to analyze biochemically and/or functionally discrete subproteomes such as phosphoproteins and ubiquitinated proteins.

Quantitative or comparative proteomics is essential for characterizing a disease proteome in relation to its normal counterpart. In traditional 2-DE–MS analyses, the staining intensity of protein spots is compared between gels to determine relative protein concentrations. A more recent method can visualize multiple protein samples on a single gel using a differential fluorescent-labeling technique (called difference gel electrophoresis), thus improving reproducibility by avoiding gel-to-gel variation. When using an LC-MS/MS approach, the samples being compared are chemically labeled with isotopically distinct (ie, different mass) tags. This allows comparison of the relative abundance of each labeled peptide in the 2 samples by analysis of peptides identical in sequence but differing in mass (Figure 1A). This method still remains a technical challenge, however, because protein coverage is compromised by inefficient chemical labeling, limited representation of peptides in a protein because of amino acid–specific labeling chemistry, sample loss ascribable to additional purification steps, and chemical side reactions. To overcome these limitations, other LC-MS/MS approaches are also being developed that do not involve differential isotopic labeling. For instance, recent studies indicate that the number of unique peptides identified for a single protein can serve as a measure of protein abundance (Figure 1B). Although this method is semiquantitative and tends to underrepresent low-abundance proteins, it can detect 2.5- to 5-fold changes in protein abundance with high confidence and is suitable for high-throughput multisample proteome profiling. Improvements in MS instrumentation are also making it possible to quantify differentially expressed proteins on the basis of the ion volume calculated for individual samples by integrating the extracted ion chromatogram for a peptide of interest (Figure 1C and D). In this application, the high-throughput, high-protein coverage and automation capability of LC-MS/MS remain intact, facilitating its use for large-scale differential expression analysis and biomarker detection as recently accomplished with a large group of human serum samples.

In addition to being a powerful discovery tool, proteomics is also evolving into an effective diagnostic tool. A directed protein detection approach, known as selected reaction monitoring (Figure 1E), is gaining popularity because of its capability of determining the absolute amount in terms of grams or moles of a single or selected set of proteins from very complex mixtures. This is a rapid and sensitive (picograms per milliliter) method and is amenable to multiplexing so that panels of biomarkers could be evaluated for a patient in a single analysis.

RELEVANCE TO THE PRACTICE OF NEUROLOGY

Biofluids contain a proteome that may reveal protein signatures likely to provide “real-time” biomarkers for diagnosis and therapeutic decision making. Unlike brain tissue, which is often only available from postmortem cases, biofluids such as serum/plasma and cerebrospinal fluid (CSF) are readily extracted from living patients. The ability of proteomic methods to extract diagnostic and prognostic information from biofluids has the potential to revolutionize the delivery of care to patients with neurological diseases. This may occur because proteomic analysis of biofluids is likely to enable both the generation of more precise diagnoses and improved means of monitoring patient responses to therapies. Both serum and CSF are a rich source of proteomic information. Using modern proteomic methods, more than 4000 proteins have been identified in serum and 1500 unique proteins have been found in CSF.

Several studies using state-of-the-art proteomic methods have already identified differences in the composition of CSF proteins for Alzheimer disease, dementia with Lewy bodies, Parkinson disease, and amyotrophic lateral sclerosis. Recent studies have also identified new proteins in CSF from patients with multiple sclerosis that were not previously recognized in human CSF and thus may constitute multiple sclerosis–specific protein markers.

Obviously, these and other future advances could provide an opportunity to refine diagnostic methods for a host of neurological disorders, as outlined in Figure 2. For example, it can be difficult to distinguish patients with Alzheimer disease from those with frontotemporal dementia at an early stage. Obtaining an accurate early diagnosis may become feasible with the advances made by proteomics, facilitating the application of disease-specific therapies. In periodic neurological diseases such as epilepsy, the identification of specific proteomic changes that might correlate with increased risk for future seizures, failure of monotherapy, or response to a particular class of antiepileptic medication may allow epileptologists to more effectively tailor medical therapy. Additionally, CSF proteomic profiles may have predictive value in identifying patients likely to fail medical therapy who thereby might be referred for earlier surgical evaluation. Multiple sclerosis is another periodic neurological disease that would benefit from CSF proteome profiling. Currently, multiple sclerosis treatment decisions are largely made based on functional criteria. Magnetic resonance imaging is the only means of obtaining objective information indicative of disease activity, but it may not be a reliable predictor of which patients will develop a relapse in the near future. If proteins are identified that strongly correlate with active disease or that can predict the future course of disease progression, it would significantly advance multiple sclerosis
Figure 1. Schematic representation of liquid chromatography tandem mass spectrometry (LC-MS/MS)–based quantification methods. Common to the 3 major methods (A-C), extracted protein samples are first enzyme digested, typically with trypsin, in the sample preparation step. The isotope-coded affinity tag (ICAT) labeling method (A) also involves a cysteine-targeted chemical labeling step prior to enzyme digestion. Since the 2 samples (proteome A and B) are combined following chemical labeling and thereafter handled in an identical manner throughout the LC-MS/MS analysis, the ability to report a quantitative difference is superior to the other 2 methods. In all of the listed methods, multiple mass spectrometry (MS) scans are obtained sequentially as peptides are separated by liquid chromatography and are sent online to the first MS step. Detected peaks are then fragmented and their masses analyzed for peptide identification (amino acid sequence determination). Using the ICAT method (A), quantification is performed by comparing the height or area of paired mass peaks (blue cones) that differ in mass only by the presence of the tag. Note that the mass peak of 1 peptide species appears as a group of peaks because of the presence of naturally occurring stable isotopic carbon species. In the peptide abundance index method (B), quantification is based on the number of unique peptides identified across MS scans relative to the number of possible peptides per protein (based on the unique trypsin cleavage pattern). In the third approach, the extracted ion chromatogram method (C), the mass peak corresponding to a peptide of interest is identified and integrated across multiple MS scans containing a signal for the peptide as shown in D (shaded area), which can then be compared among samples run separately. With the increasingly improved mass accuracy of mass spectrometers, identification of a peptide can solely rely on the mass of a detected peak, eliminating the need for MS/MS analysis. Integration can also be applied to the ICAT method for more accurate quantification. These 3 major methods of quantification differ in their accuracy of relative quantification, protein coverage, and throughput as denoted by ++ (good) to +++ (excellent) in the Figure. Lastly, a targeted detection approach under development, known as selected reaction monitoring (E), has the potential to facilitate sensitive quantification of select proteins in complex biological samples. A peptide carefully selected from a target protein for reproducible MS behavior is synthesized with, for instance, carbon 13 ($^{13}$C) and spiked into a sample to be analyzed. The synthetic peptide serves as an internal standard that is chemically identical to the native peptide but has a greater mass than the native peptide by a known amount (X Da). The peptide mimic serves as a marker peptide to identify the location of the native peptide during MS analysis and an internal standard for quantification achieved by mass peak integration as shown in D. $^{12}$C indicates 9 carbon 12 atoms; $^{13}$C, 9 carbon 13 atoms; m/z, mass-to-charge ratio.
therapy. For instance, potentially toxic treatments such as natalizumab and mitoxantrone hydrochloride might be indicated only for those patients whose proteomic profile predicts rapid disease progression.

Despite the many potential benefits associated with proteomics, the feasibility of implementing current proteomic approaches in a clinical setting still appears very limited. The global approach remains technically demanding and labor intensive and requires significant resources in terms of equipment and personnel. Therefore, the field would benefit from a concerted collaboration between experts in the field of proteomics and health care professionals treating patients with neurological diseases, focusing on the establishment of institutional neuroproteomic centers. These centers would ideally facilitate the accrual and banking of patient samples for eventual proteomic analysis along with archiving extensive clinical information needed to rigorously characterize and validate the significance of protein profiles or specific proteins as putative biomarkers. These centers would then be situated to perform a secondary prospective analysis needed to further evaluate and validate specific biomarkers as they are identified.

RELEVANCE TO THE STUDY OF NEUROSCIENCE

The comprehensive nature of proteomic analysis can provide a unique perspective of normal and diseased proteomes, which may lead to unexpected findings and a new understanding of both normal biological and disease processes. Proteomic approaches have already enhanced our understanding of the mechanisms of neurotransmission. While individual neurotransmitter receptors can be expressed in cellular systems, such as Xenopus oocytes that allow for exquisitely detailed understanding of how the structure of these proteins affect their ability to generate current in response to ligands, these methods do not reveal the complete picture of how receptor structure/function is actually modulated in the postsynaptic cell. Proteomic analyses have revealed that the N-methyl-D-aspartate–type glutamate receptor is part of a large protein complex that includes close to 200 proteins either specifically or commonly associated with the NR1 and NR2B subunits while more than 1000 proteins are present in the entire postsynaptic proteome.\(^9\) The proteomic data collected thus far have facilitated the creation of putative protein × protein interaction maps and signaling networks that were previously unattainable by other approaches.\(^{25}\)

Defining the proteome of specific neural cell types, tissues, and neuroanatomical regions will provide a body of information of unprecedented size and value. Future proteomic experiments that characterize state-dependent changes in the proteome or protein complexes will be compared with these data sets in a manner that is similar to the use of neuroanatomical atlases that helped define neuroscience in the previous century. Take peripheral nerves as an example. We do not yet know all of the proteins required for normal function of intact peripheral nerve tissue. Motor neurons, sensory neurons, or Schwann cells cultured in isolation from other cell types normally present in peripheral nerves likely have proteomic profiles that are unique to the isolated state. Comparison of the profiles from each cell type with that of the whole explanted nerve could lead to the identification of specific proteins expressed only when the appropriate cellular and matrix associations are extant. These newly identified proteins could turn out to have extremely important roles in normal peripheral nerve function and could even be important factors in the large variety of peripheral nerve diseases for which we currently have no identified pathogenetic mechanism.

Characterizing the proteome of subcellular structures such as specific organelles will also provide important new information about nervous system function. Mitochondria represent one example of a subproteome that has been extensively analyzed by proteomics. Thus far, 600 to 700 mitochondrial proteins have been identified in mammals, representing approximately 50% of estimated mitochondrial proteins.
Interestingly, the mitochondrial proteome in the brain is substantially different from that in other tissues. Because alterations in mitochondrial function are commonly seen in aging and neurodegenerative diseases, results of these proteomic studies should help in understanding the normal function of mitochondria and their contribution to neuronal dysfunction following stress. The selective vulnerability of specific neurons to a certain stress may stem from heterogeneity of their mitochondrial proteome. This heterogeneity may determine, for example, what types of proteins bind to mitochondria and play a role in localizing mitochondrion within axons and dendrites and what proteins translocate to the mitochondria in response to cellular stress and compromise mitochondrial integrity. Comparative studies of mitochondria and other subcellular structures may also provide a unique perspective of the types of communication taking place between organelles.

Protein aggregates and inclusions represent protein complexes directly implicated in neuropathological processes and are relevant targets for proteomic analysis. In one such study using a cell culture model of Parkinson disease, more than 250 proteins were identified in α-synuclein protein complexes, with about 30 proteins showing altered levels of expression in response to rotenone treatment.23 A complete list of the proteins contained in protein aggregates, such as Lewy bodies, Hirano bodies, amyloid plaques, neurofibrillary tangles, and polyglutamine protein aggregates, may help elucidate the roles played by the formation of these aggregates and their components in disease pathogenesis. Since proteomic identification of the constituents of protein complexes isolated by immunofinity or, more recently, by tandem affinity purification does not require prior knowledge of their identity, this proteomic approach is very powerful and is uniformly applicable to many questions involving protein−protein interactions currently being addressed in the neurosciences.

Finally, successful proteomic analyses, whether related to the identification of biomarkers for a disease or the characterization of protein complexes, yield an enormous amount of data. To be of genuine benefit, these proteomic results must be validated using additional approaches and models. The use of in vivo and in vitro models of specific neurological diseases, when available, allows one to reproducibly obtain tissues or cells at specific points along the course of disease development, offering investigators a greater degree of control over the type and state of the tissue or cells being analyzed. The combined use of in vivo and in vitro models and patient tissues or biofluids may provide significant insight into disease-related proteomic changes that will hopefully translate into more effective and specific methods for diagnosis and treatment of neurological disorders.

Accepted for Publication: July 31, 2006.

Correspondence: Richard S. Morris, PhD, Department of Neurological Surgery, University of Washington School of Medicine, Box 356470, Seattle, WA 98195-6470 (yael@u.washington.edu).

Author Contributions: Study concept and design: Kinoshita, Uo, Jayadev, Garden, Conrads, Veenstra, and Morrison. Analysis and interpretation of data: Jayadev. Drafting of the manuscript: Kinoshita, Jayadev, Garden, and Morrison. Critical revision of the manuscript for important intellectual content: Uo, Jayadev, Garden, Conrads, Veenstra, and Morrison. Obtained funding: Morrison. Administrative, technical, and material support: Conrads. Study supervision: Garden, Veenstra, and Morrison.

Financial Disclosure: None reported.

REFERENCES


©2006 American Medical Association. All rights reserved.