Proteomic Analysis in the Neurosciences*

Richard S. Morrison‡§, Yoshito Kinoshita‡, Mark D. Johnson‡, Takuma Uo‡, Joseph T. Ho‡, Joshua K. McBee¶, Thomas P. Conrads||, and Timothy D. Veenstra||

Proteomics is a field of study directed toward providing a comprehensive view of the characteristics and activity of every cellular protein. Rapid innovations in the core technologies required to characterize proteins on a global scale are poised to bring about a comprehensive understanding of how dynamic changes in protein expression, post-translational modification, and function affect complex signaling and regulatory networks. These advances have significant implications for understanding the multitude of pathways that govern behavior and cognition and the response of the nervous system to injury and disease. *Molecular & Cellular Proteomics* 1:553–560, 2002.

Recent advances in molecular biology, instrument technology, and bioinformatics are making it possible to simultaneously analyze the entire complement of genes expressed in a particular cell or tissue. These advances have created unique opportunities in the field of medicine, where the results from gene expression studies are expected to help identify cellular alterations that are associated with disease etiology, progression, outcome, and response to therapy. These rapidly emerging technologies are also expected to result in the identification of novel therapeutic targets for a host of maladies, including infectious diseases, behavioral disorders, developmental defects, neurodegenerative diseases, aging, and cancer.

Technical advances have facilitated the characterization of the three major genetic units, the genome, the transcriptome, and the proteome. The genome describes the entire set of genes that is encoded by the DNA of an organism. The transcriptome encompasses the entire complement of mRNA transcripts transcribed from the genome of a cell. The transcriptome varies from cell to cell and fluctuates in response to numerous physiological signals including developmental cues, stress, changes in the extracellular milieu, and disease. The proteome describes the entire complement of proteins expressed by a cell at a single point in time. Proteomic investigations also aim to determine protein localization, modifications, interactions, and ultimately protein function. Because the function of a gene is dependent on the activity of its translated protein, there has been significant impetus to develop methods that will enable high throughput analysis of cellular proteomes. Taken together, the study of the genome, transcriptome, and proteome provides complementary insights into a host of biological processes and affords a greater understanding of the regulation of these processes.

PROTEOMICS FILLS A NICHEN

The rapidly evolving field of proteomics is directed toward providing a comprehensive view of the characteristics and activity of every cellular protein. The proteome is clearly more complicated than the genome as a single gene can encode multiple forms of a protein. This variable expression can result from the following: 1) alternative splicing of the mRNA transcript, 2) the use of alternate translation start or stop sites, and 3) the occurrence of frame-shifting, during which a different set of triplet codons in the mRNA is translated. The net result of these activities is the generation of a proteome that contains many proteins derived from shared or overlapping genomic sequences.

The importance of studying cells at the proteome level is further underscored by the difficulty in predicting protein characteristics from genomic sequence data alone. These characteristics include post-translational modifications, subcellular distribution, stability, biomolecular interactions, and function. In contrast to DNA and RNA, proteins can be modified by phosphorylation, glycosylation, acetylation, nitrosation, poly(ADP-ribosylation), ubiquitination, farnesylation, sulfation, linkage to glycosylphosphatidylinositol anchors, and SUMOylation. In total, there are about 300 different post-translational modifications that have been reported (1). These modifications can profoundly affect protein conformation, stability, localization, binding interactions, and function. Proteins are often modified at multiple sites, and it is not possible to predict from an amino acid sequence with complete certainty which sites will be modified in response to a specific set of conditions. The p53 tumor suppressor protein provides a striking example of a protein that is affected profoundly by the type and location of modifications including phosphorylation, acetylation, and SUMOylation. The phosphorylation of certain serine residues is required for p53-mediated transcription of several downstream targets associated with cell cycle arrest, including p21WAF1/Cip1 (p21) and mdm-2 (2). In contrast to these specific serine residues, phosphorylation at other serine residues (serine 46) regulates the transcriptional activation of apoptosis (3–5). Post-translational addition of the ubiquitin-
like protein, SUMO-1, appears to be important for protein targeting and protein-protein interactions. For example, targeting of the RanGAP1 protein to the nuclear pore is absolutely dependent upon the SUMO-1 modification (6). Although it is possible to analyze a genetic sequence for the presence of putative consensus sites for various post-translational modifications, the mere presence of such sites does not indicate whether they are utilized, under what circumstances they are utilized, or if they are utilized simultaneously.

Another powerful impetus for moving beyond the transcriptome is the demonstration by several researchers that protein levels do not faithfully correlate with mRNA levels (7–9). An analysis of 106 genes in the yeast *Saccharomyces cerevisiae* demonstrated that the levels of protein expression attributed to mRNA species of equal abundance could vary by as much as 30-fold. Conversely, the mRNA levels for proteins that were expressed at comparable levels varied as much as 20-fold. Experience from our own laboratory with cDNA microarray analysis yielded similar results. Analysis of DNA damage-induced cell death in postnatal cortical neurons demonstrated that only 48% of mRNA/protein pairs displayed a change in mRNA abundance that was in the same general direction as the measured changes in protein levels.1

**DIVERSE PROTEOMIC APPROACHES ABOUND**

Proteome analysis is generally driven by a two-stage process, protein separation and mass spectral analysis. Multidimensional separations are required to minimize the number of peptides being introduced to the mass spectrometer for analysis at any given time. A reduction in the number of peptides within any given spectrum will increase the overall dynamic range of the MS2 measurements, thereby optimizing the detection of lower abundance peptides, even using conventional instrumentation. A second approach advocates combining a single dimensional separation with advanced mass spectrometric instrumentation. Although a high resolution separation is still necessary, the gains in dynamic range and sensitivity needed to identify low abundance proteins are ultimately provided by the MS instrumentation. The MS developments focus primarily on methods to introduce as many peptides into the instrument as possible, as well as ways to manipulate the ion population so that lower abundance species can be measured exclusively within a complex sample. At this point in proteomic technology development it seems likely that the broadest proteome coverage will come from a combination of multidimensional fractionation and advanced MS instrumentation.

Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) is currently considered the cornerstone of protein fractionation in proteomics. 2D-PAGE produces high resolution protein separations, based on the molecular mass and pI of the protein, resulting in the display of potentially thousands of protein spots (10). The large number of protein spots often conveys the impression that this technology provides a relatively complete view of the proteome. Unfortunately, several factors limit the resolving power of 2D-PAGE. For example, hydrophobic and membrane proteins exhibit incomplete solubility in the 2D-PAGE system, and very large or small and highly basic or acidic proteins are difficult to resolve. In addition, a single gene can be represented by multiple spots on a gel, and conversely, a single spot can be composed of several different gene products (11). The complexity of spot patterns arises from post-translational modifications, alternative splicing, protein degradation, and artifacts of the 2D-PAGE system. Methods for identifying protein spots can also be problematic in that they are slow, expensive, and often insensitive. Even the combination of 2D-PAGE and MS only detects the most abundant proteins (12), indicating that the 2D-PAGE approach does not provide sufficient coverage of the proteome.

To circumvent the use of 2D-PAGE separations, many groups are developing solution-phase, multidimensional fractionation techniques and more powerful mass spectrometric technologies to enhance the identification of low abundance proteins. For example, a method has been developed to analyze protein digests using a multidimensional capillary liquid chromatography (LC) separation in which tryptic digests of proteome samples are separated by strong cation exchange chromatography in the first dimension followed by reversed-phase LC in the second. This fractionation system, referred to as MudPIT, is coupled directly on-line with MS and is capable of identifying thousands of proteins in a single automated analysis coupled with tandem mass spectrometry (13, 14). Another novel method currently being evaluated involves the analysis of global proteolytic digests using a single dimension of ultrahigh resolution capillary reversed-phase LC, in combination with powerful Fourier transform ion cyclotron resonance (FTICR) MS (15). A key aspect of using FTICR in proteomics is its ability to routinely obtain very high mass measurement accuracy (MMA) for the large numbers of proteins or proteolytic peptides delivered to the mass spectrometer by on-line separations. FTICR instrumentation can obtain an MMA of ~0.1 ppm (although much more accurate measurements have been demonstrated for specialized applications), which is higher than the 5 to 20 ppm MMA achievable with the best conventional MS technologies and the 100–500 ppm accuracy of the most widely used quadrupole and ion trap MS technologies. The high MMA achievable with FTICR may allow proteins to be identified uniquely based solely on the mass of a single proteolytic fragment, obviating the need for time-consuming tandem MS experiments. The proteolytic fragment can then act as a biomarker or accurate mass tag to

---


2 The abbreviations used are: MS, mass spectrometry; FTICR, Fourier transform ion cyclotron resonance; ICAT, isotope-coded affinity tag; LC, liquid chromatography; MMA, mass measurement accuracy; 2D, two-dimensional; HPLC, high pressure liquid chromatography; AD, Alzheimer’s disease; GSK, glycogen synthase kinase; EDT, ethanedithiol; SUMO, small ubiquitin modification.
Proteomic methods are also being developed to facilitate the quantitative analysis of protein expression in complex mixtures of peptides. Aebersold and co-workers (17) have developed a novel isotope-coded affinity tag (ICAT) strategy that permits the stable-isotope labeling of cysteine residues in proteins, thus facilitating a quantitative global analysis of differences in protein expression (Fig. 1). In the ICAT approach, proteins are reacted with a thiol-specific reagent containing a biotin functionality group. The thiol-specific group allows for the chemical modification of reduced cysteine residues, whereas the biotin group is used in combination with immobilized avidin to isolate these modified peptides. This extraction of the ICAT-modified peptides following trypsin digestion significantly reduces the complexity of the mixture and facilitates the identification and quantitation of lower abundance proteins. Two isotopically distinct versions of the ICAT reagent are available, a light isotopic version and a heavy isotopic version in which eight protons within the linker region have been substituted with eight deuterons. The ICAT approach allows for a significant reduction in the complexity of the polypeptide mixtures while maintaining broad proteome coverage, as well as providing for proteome-wide precise quantitation of protein expression levels when combined with advanced MS technologies. Although metabolic labeling strategies offer the highest precision for quantitating protein abundances and the broadest proteome coverage, post-isolation isotopic labeling is broadly applicable to proteins extracted from any conceivable source (e.g. tissue samples). This provides an opportunity to obtain quantitative global measurements of proteomes in the context of a biological or pathological process.

Although the ICAT approach shows great promise for determining relative levels of protein expression, delineation of protein function solely by changes in abundance will be limited, because a multitude of protein activities are modulated by post-translational modifications. One of the most important post-translational protein modifications used to modulate protein activity and propagate signals within complex cellular pathways is phosphorylation (18–20). Studies estimate that as many as one-third of all cellular proteins in mammalian cells are phosphorylated (20). Cellular processes such as cell cycle progression, cellular differentiation, malignant transformation, development, and neurochemical activity are all regulated by changes in the state of protein phosphorylation. Therefore, the ability to broadly identify changes in the phosphorylation state of proteins may lead to a better understanding of how specific protein activities are related to complex signaling networks within cells and tissues.

The predominant method used to study changes in protein phosphorylation is by metabolically labeling proteins with $^{32}$P inorganic phosphate ($^{32}$P). To measure relative changes in the degree of phosphorylation, $^{32}$P-labeled proteomes are resolved using 2D-PAGE, and the relative spot intensities are compared (21, 22). The use of $^{32}$P, to label proteins does not lend itself to high throughput proteome-wide analysis because of the problems with handling radioactive compounds and the associated contamination of instrumentation. More rapid and comprehensive methods are needed to identify phosphorylated proteins and quantitate the extent of phosphorylation.

In its simplest form, MS can be used to provide an accurate mass measurement of an intact phosphorylated protein. Comparing this mass to the calculated mass of the unmodified protein or the mass of the protein after phosphatase treatment allows the number of bound phosphate groups to be calculated (23). Unfortunately, analysis of the intact protein by this method does not provide any information related to the
Proteomic Analysis in the Neurosciences

specific site of phosphorylation, a key piece of information that can directly affect the function of the protein. To identify specific phosphorylated residues requires analysis of the protein at the peptide level. Peptides are generated by enzymatic or chemical digestion of the intact protein, which are then analyzed by either MS or tandem MS (24). Although MS measurements can confirm the presence of a phosphate group on a peptide, tandem MS is still necessary to establish the specific site of phosphorylation when two or more residues are phosphorylated.

One of the major difficulties in the analysis of phosphopeptides is their relatively low abundance when compared with non-phosphorylated peptides within the sample (25). Because the presence of other peptides can suppress the ability to detect phosphopeptides by MS, phosphopeptide detection can be enhanced using methods that reduce the amount of non-phosphorylated peptides within the mixture. Generating a mixture enriched for phosphopeptides not only increases the ability to detect this class of peptides but also aids in the downstream analysis, because the identification of a greater percentage of phosphopeptides can be anticipated. Several strategies have been developed to enrich the sample for phosphorylated peptides or phosphoproteins before MS analysis. Phospho-specific antibodies and immobilized metal-affinity columns have been widely employed, but they typically result in isolation of non-phosphorylated species, along with the phosphopeptides of interest. Recently, however, Hunt and co-workers (26) reported a multidimensional separation technology that greatly improves the specificity of immobilized metal-affinity columns. Immobilized metal-affinity columns were used to prepare an enriched mixture of phosphopeptides, which were subsequently analyzed by nano-flow HPLC/MS. They detected a total of 216 peptide sequences defining 383 sites of phosphorylation from an analysis of a whole-cell lysate from *S. cerevisiae*.

Two new methods have been developed that facilitate the specific enrichment and quantitation of phosphopeptides (27–30). Both methods utilize stable isotopes to differentially label the samples being compared and subsequently employ MS analysis for the identification and quantitation of the enriched phosphopeptide mixture (Fig. 2 and Fig. 3). An important advantage of the isotopic chemical modification strategies is that the modification remains attached to the residue during tandem MS fragmentation of the peptide. During tandem MS the intact peptide is subjected to collision induced dissociation, where the ion selected collides with an inert gas (nitrogen or helium, for example) that causes it to fragment into smaller ions. The MS spectrum of these fragment ions typically provides partial sequence information that can be used in conjunction with commercially available computer algorithms to identify the peptide. In a typical MS experiment, however, the phosphate group dissociates from the phosphorylated residues during the tandem MS analysis (indeed, often in MS analyses, as well) preventing the site-specific assignment of the phosphate modification. The isotope labeling strategy allows the exact phosphorylation site to be determined by tandem MS.

**Fig. 2. Isotope-affinity strategy for isolation and quantification of phosphopeptides.** Top, proteins containing phosphoserine (X = H) or phosphothreonine (X = CH$_3$) residues are modified with reagents containing both an isotopically labeled linker and biotin group. After proteolytic digestion, these modified peptides are isolated using immobilized avidin-affinity chromatography. Light (L = H, EDT-D$_0$) and heavy isotopic versions (L = D, EDT-D$_4$) of 1,2-EDT are used to quantitate the relative phosphorylation state of phosphopeptides extracted from two different samples. Bottom, mass spectrum of EDT-D$_0$/D$_4$-labeled β-casein peptide. The enriched mixture of biotinylated phosphopeptides was analyzed by capillary reversed-phase liquid chromatography coupled directly on-line to a hybrid quadrupole-time of flight mass spectrometer. The [M+2H]$^{2+}$ ion pair corresponds to the mass of the derivatized β-casein phosphopeptide FQS*EEQQUOTEDELQDK where *S* has the modified side chain -CH$_2$-SCL$_2$-S-acetyl-PEO-biotin, and L is either H (EDT-D$_0$) or D (EDT-D$_4$). The isotopic envelopes representing the peptide labeled with EDT-D$_0$ or EDT-D$_4$ is labeled in the mass spectrum.

**PROTEOMICS ANALYSIS OF NERVOUS SYSTEM PHYSIOLOGY AND PATHOLOGY**

The proteomic methods described above can be used to produce global profiles of protein expression. Thus, a proteome can be defined for normal tissues/cells or tissues/cells subjected to a specific treatment, injury, or disease. Constructing global maps of protein expression is currently the
most widely employed application of proteomics akin to using DNA microarray analysis to define comprehensive patterns of mRNA expression. The earliest studies to apply this technology are beginning to provide important information about the diversity of proteins expressed in the nervous system and how they vary in response to injury and disease (Fig. 3). Two-dimensional gel electrophoresis, in combination with several forms of MS, was used to initiate a proteome map of the adult cerebellum from mouse (31), rat (32), and pig (33). In one study, protein profiles from the developing and adult rat cerebellum were analyzed by means of 2D-PAGE and MS (32). Over 3000 spots were resolved on the gels from adult cerebellum with 67 of these being assigned identities and used as landmarks for comparison with profiles obtained from developing cerebellum (postnatal days 0, 3, 7, 14, and 21). Although most proteins showed an increase in abundance as the cerebellum matured, 42 spots appeared to be exclusively expressed or highly abundant in immature cerebellum. Twenty-nine of the latter were identified by peptide mass fingerprinting and included proteins of unknown function and those with defined roles in the development of the nervous system. Two-dimensional protein maps have been constructed for the whole rat (11, 34) and human brains (35) to analyze systematically the entire brain proteome. In these studies more than 200 discrete proteins were identified. There was a significant degree of overlap in the proteins identified between the two studies suggesting that the techniques are reliable and reproducible at least with respect to the more abundant protein species. Proteins from all conceivable functional categories were identified including structural proteins, enzymes involved in energy metabolism, protein synthesis, protein degradation, RNA transcription, oxidative stress response, and signal transduction, as well as synaptic proteins. Because these maps were constructed from whole brain, proteins were derived from neurons, astrocytes, oligodendrocytes, microglia, and blood cells. The use of purified cell cultures, flow
cytometry, immunopanning, laser capture microdissection, and other selection techniques should reduce the complexity of protein mixtures and facilitate the identification of cell typespecific and lower abundance proteins. Nevertheless, the construction of comprehensive brain proteome maps should prove useful as reference databases to study changes in protein levels associated with development, aging, behavioral abnormalities, or various disorders afflicting the nervous system such as Alzheimer’s and Huntington’s diseases.

Proteome analysis is gradually being applied to human brain samples from patients with neurodegenerative diseases or to the brains of transgenic mice that model human neurodegenerative diseases. Comparative proteome analysis was performed in post-mortem brain tissues from patients with Alzheimer’s disease (AD) and compared with profiles from age-matched, non-demented control brain tissue (36). Proteins were resolved by 2D-PAGE and identified by NH2-terminal sequence analysis. Thirty-seven proteins that differed significantly between AD tissue and control tissue were identified previously in the pathogenesis of AD.

In a related study, a proteomics approach was used to identify proteins that might play a role in the cellular abnormalities caused by overexpression of human tau (37). Abnormal tau expression has been implicated in the pathogenesis of several human neurodegenerative disorders including AD, Down’s syndrome, Pick’s disease, and others (38–40). Proteins were resolved by 2D-PAGE, and 34 proteins whose expression levels in wild-type and tau transgenic mice differed at least 1.5-fold were identified by electrospray ionization tandem MS. Once again the identified proteins could be classified into discrete functional categories including carbohydrate metabolism, lipid transport, stress response, and neurotransmission. A number of proteins identified in this study, such as α-crystallin, superoxide dismutase, glyceraldehyde phosphate dehydrogenase, and dihydroorotate-related protein have been implicated previously in the pathogenesis of AD.

There is significant evidence that acutely injured neurons die by an active mechanism of cell death. This process involves the activation of discrete signaling pathways that ultimately compromise mitochondrial structure, energy metabolism, and nuclear integrity. An injury model involving the systemic injection of kainic acid, a potent excitotoxin that produces seizures accompanied by neuronal cell loss, was evaluated for changes in brain protein levels using 2D-PAGE and matrix-assisted laser desorption ionization MS (48, 49). Protein levels were evaluated 7 days after injection, a time point corresponding to significant neuronal damage and cell loss. Although this time point is too late to visualize early biochemical changes associated with excitotoxicity-induced cell death (50), the data likely reflects important changes in protein expression associated with repair processes activated in the surviving population of neurons. Thus, decreased expression of neurofilament L and M proteins are consistent with neuronal cell loss, whereas increased expression of dihydropyrimidinase-related proteins is reflective of repair processes. Interestingly, the dihydropyrimidinase-related proteins have been implicated in the regulation of neurite outgrowth during brain development (51).

Proteomic analysis recently demonstrated changes in the semaphorin/collapsin family and dihydropyrimidinase-related proteins in fetal brain samples from patients with Down’s syndrome (52). A similar approach was also used to evaluate expression levels of actin-binding proteins in fetal Down’s syndrome brain and control cortex (53). The authors (53) noted a significant reduction of the actin-related protein complex 2/3 (Arp2/3) 20-kDa subunit and the coronin-like protein p57, which are involved in actin filament cross-linking and nucleation and capping of actin filaments. Using an in vitro model of DNA damage-induced cell death in postnatal cortical neurons, our laboratory also identified significant changes in the expression of proteins associated with neurite outgrowth (rather than neurite degeneration) by MS (54). These data collectively demonstrate that certain forms of brain injury and neurodegenerative diseases are associated with abnormal...
malties in the expression of proteins responsible for regulating the growth and maintaining the integrity of neuronal processes. Alterations in this related family of proteins, defined by proteomics analysis, and ensuing changes in neural connectivity might contribute to the decline of neuronal function and viability observed in these various neurological conditions.

SUMMARY

The development of the nervous system, its function, and its continued viability are initiated and maintained through a complex interacting network of signaling pathways and cellular interactions that can be perturbed in response to a multitude of cellular stresses. A shift in the balance of signaling pathways after stress or in response to pathology can have drastic consequences for the development or the normal function of the nervous system. Characterizing the myriad activities necessary to support these complex processes is a daunting task, but technical developments in the field of proteomics are poised to generate advances in our understanding of protein expression, function, and organization in complex signaling and regulatory networks. Improvements in MS instrumentation, the implementation of protein arrays, and the development of robust informatics software are providing sensitive, high throughput technologies for large scale identification and quantitation of protein expression, protein modifications, subcellular localization, protein-protein interactions, and protein function. These advances have significant implications for understanding how cellular proteomes are regulated in the nervous system in health and disease.

* This work was supported in part by National Institutes of Health Grants NS35533, NS42123, and NS42699 (to R. S. M.), National Institutes of Health Vision Training Grant EY07031 (to J. K. M.), and grants from the NCI, National Institutes of Health under Contract NO1-CO-12400 (to T. D. V.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ To whom correspondence should be addressed: Dept. of Neurological Surgery, University of Washington School of Medicine, Box 356470, Seattle, WA 98195-6470. Tel.: 206-543-9654; Fax: 206-543-8315; E-mail: yaell@u.washington.edu.

REFERENCES

phorylated proteins as a tool for probing the phosphoproteome. Nat. Biotechnol. 19, 379–382