Use of Quantitative Mass Spectrometry Analysis in Kidney Research

Timothy D. Cummins, MS, and David W. Powell, PhD

Summary: An important component of nephrology research is the discovery of novel proteins that control cellular and molecular events that contribute to normal kidney cell biology and disease. Identifying perturbation of normal cellular protein expression and interactions within signaling networks is critical for understanding these regulatory events. Methods that couple 2-dimensional capillary liquid chromatography and tandem mass spectrometry (2D-LC-MS/MS) analysis have greatly facilitated this discovery science. Coupling 2D-LC-MS/MS analysis with automated genome-assisted spectra interpretation allows a direct, high-throughput, and high-sensitivity identification of hundreds to thousands of individual proteins from targeted complex biological samples. The systematic qualitative and quantitative comparison of experimental/disease conditions and appropriate controls allow protein function or disease states to be modeled. This review discusses the different purification and quantitative strategies that have been developed and used in combination with 2D-LC-MS/MS and computational analysis to define regulatory events in kidney biology and disease.

Keywords: *Mass spectrometry, liquid chromatography, 2D-LC-MS/MS, protein interactions, quantitative analysis*

An important goal of nephrology research is the discovery of novel molecular targets for the effective diagnosis and treatment of kidney disease. Proteomic research is playing an increasingly important role in this discovery process. Advances in mass spectrometry–based proteomic technologies have lead to the high-throughput identification of proteins that control critical molecular and cellular functions.[1-3] These discoveries have been greatly facilitated by an approach that combines 2-dimensional capillary-scale liquid chromatography (2D-LC) peptide separation with tandem mass spectrometry (MS/MS).[2,4] This 2D-LC-MS/MS methodology allows for a direct, unbiased, and high-sensitivity identification of hundreds to thousands of individual proteins from virtually any type of biomedical sample[3-5] and thus has become a preferred method in proteomics. The workflow for this approach is shown in Figure 1. Selectively purified or enriched protein mixtures are enzyme digested directly in solution and the resulting peptides are then fractionated by 2D-LC and eluted directly into the mass spectrometer for MS/MS or peptide fragmentation analysis. Proteins are then identified using genome-assisted computational analysis of the acquired spectral data.[6,7] Different informatic platforms have been designed to assist functional predictions from these large protein datasets.[8]

Because these studies are at the front end of the discovery process, the validity of protein identifications and their relevance to the biological question is still paramount. A key component in acquiring this type of high-validity data is the development of computational algorithms that statistically measure the probability of proteins identified from the mass spectral data.[9,10] Rigorous computational algorithms and

Department of Medicine and Department of Biochemistry and Molecular Biology, School of Medicine, University of Louisville, Louisville, KY.

Supported by funding from the office of Science Financial Assistance Programs, US Department of Energy, and the National Institutes of Health grant DK176743.

Address reprint requests to David W. Powell, Department of Medicine and Biochemistry and Molecular Biology, School of Medicine, University of Louisville, Louisville, KY 40202. E-mail: david.powell@louisville.edu

0270-9295/07/$ - see front matter

 doi:10.1016/j.semnephrol.2007.09.008

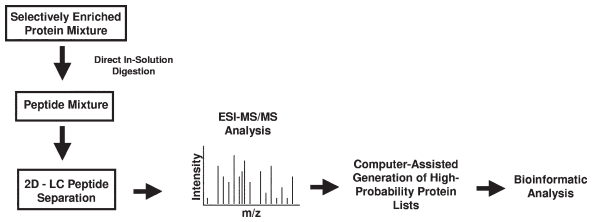

Figure 1. 2D-LC-MS/MS analysis of complex biological samples. Selectively purified or enriched protein mixtures are digested with a proteolytic enzyme of choice directly in solution. Digestions with trypsin or dual digestions with trypsin/chymotrypsin or trypsin/Lys-C typically are performed. To achieve a high level of separation and sensitivity, the resulting peptides are fractionated using 2D capillary-scale liquid chromatography. The most commonly used chemistries are strong cation exchange for the first dimension and reversed-phase for the second dimension. As the peptides elute into the mass spectrometer they are ionized and selected for MS/MS analysis. This generates spectra of ions resulting from a series of collision-induced peptide bond fragmentations. High-probability peptide and protein assignment is accomplished using computational algorithm-assisted database correlations (SequestSorcerer, http://www.sagenresearch.com/products.html) and filtering (ProteinProphet, http://tools.proteomecenter.org/software.php). These lists of proteins are submitted to bioinformatic platforms (Ingenuity Systems, http://www.ingenuity.com or DAVID Bioinformatics Database, http://david.abcc.ncifcrf.gov/content.jsp?file=about_us.html) to assist predictions of regulatory models.

statistical methodology allows for the reduction of false-positive rates and, therefore, improved identifications. Further, central to defining a protein's contribution to specific cellular and disease processes is the ability to accurately assess the relative quantity of individual proteins in the targeted biological samples. Different strategies for quantitative assessment of protein abundance from mass spectrometry data have been developed and tested for quantitative profiling of a variety of complex biological samples.[11-15] In addition, computational programs for the statistical comparison of multiple experimental/disease samples and controls are being developed.[8,9,16,17] This will allow modeling of proteomics data to identify potential targets for molecular intervention for specific diseases. This article describes different types of sample purification strategies and general components of the 2D-LC-MS/MS approach and then discusses how these methodologies are being applied to the discovery of proteins that regulate critical cellular processes in kidney biology and disease.

DEFINING COMPONENTS OF SELECTIVELY ISOLATED PROTEIN COMPLEXES WITH 2D-LC-MS/MS ANALYSIS

Key to defining molecular events that regulate critical cellular functions is the elucidation of protein networks. The basic reason for defining protein networks is that identification of proteins that co-purify with a particular target protein or complex will help define events that regulate a known biological process of the targeted proteins or cellular system. For example, targeted purification of a transcription activator/repressor with known involvement in renal fibrosis may lead to a better understanding of the disease mechanism. Another more global example is characterization of plasma membrane proteins in podocytes. The systematic purification of targeted protein complexes coupled with mass spectrometric identification has allowed regulatory protein maps of different cell types or signaling pathways to be constructed.[18-25] These maps have greatly aided in our understanding of the function and regulation of proteins in the cell. This section of the article describes some of the strategies that have been developed and applied for the purification and enrichment of targeted protein complexes from cell lines and tissues (Fig. 2).

Native protein immunoaffinity purification and epitope/fusion-tag purification are 2 general types of approaches used for affinity purification of targeted proteins.[20] The immunoaffinity purification method uses antibodies to native targets, enabling one to universally isolate protein complexes from cellular extracts, eliminating issues associated with genetic overexpression and extraneous fusion tag-mediated effects. This method is advantageous especially in studies of tissue or primary cells, where genetic manipulation is limited. However, all antibodies have some degree of cross-reactivity to proteins other than the target protein, which results in false-positive identifications. Thus, when feasible, high-affinity epitope tagging offers an effective alternative.

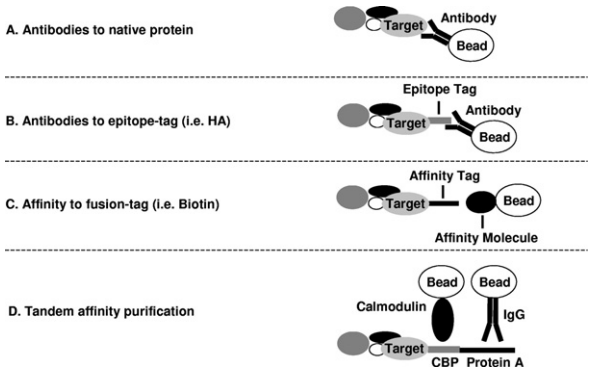

Figure 2. Approaches used to purify targeted protein complexes from cell or tissue extracts. These approaches use affinity molecules coupled to a beaded matrix. This allows for isolation of targeted protein complexes with centrifugation or gravity flow chromatography. After the initial protein recovery, the beads are washed to minimize recovery and identification of nonspecific proteins. (A) This approach uses polyclonal antibodies designed against a specific target protein. (B) For this approach the target protein is expressed with a fusion epitope tag attached to its C- or N-terminus and recovered using an antibody against the tag. (C) This method uses non–immune-based affinity interactions, such as glutathione S-transferase/glutathione or biotin/streptavidin. (D) This is a 2-step affinity-purification approach. The target protein is expressed with 2 high-affinity tags (calmodulin binding peptide [CBP] and a protein A sequence) attached to its C- or N-terminus. A specific protease cleavage site is placed strategically between the 2 different affinity tags. This allows elution or release of the targeted complex from the first affinity (immunoglobulin G [IgG]) step. Then the complex is recovered in a second step using a calmodulin matrix.

An innovation in epitope tagging that was developed initially in yeast uses a tandem affinity purification (TAP) methodology (Fig. 2D).[19] By using 2 different extremely high-affinity epitope tags (protein A and calmodulin binding sequences), nonspecific interactions are minimized dramatically, allowing selective isolation of relatively pure target complexes. The use of universal affinity reagents and methodology also eliminates the need for specialized reagents and optimization for different targets and offers an effective comparative experimental control for assessing target-specific interactions. The TAP approach has been used routinely in combination with 2D-LC-MS/MS to discover novel protein–protein interactions in selective protein complexes isolated from model organisms, as well as mammalian cells.[3,18,22]

An important facet of establishing an effective high-throughput means to characterize regulatory protein networks is that it provides an expedient and systematic approach to elucidate specific molecular mechanisms regulating the structure and function of the targeted complex (Fig. 3). A 2D-LC-MS/MS analysis is used to compare the composition of the targeted complexes isolated from cells or tissues from experimental/disease conditions versus appropriate controls. For example, a model for chronic kidney disease could compare the changing composition of protein complexes resulting from exposure of renal cells to different pro-inflammatory cytokines.[26] An example of examining a more direct molecular process is determining the effects of deleting or mutating a component(s) of the target complex.[3,27] Quantitative comparison of experimental/disease purification with controls is facilitated using isotope-labeling or label-free approaches. These techniques are discussed in greater detail in a later section.

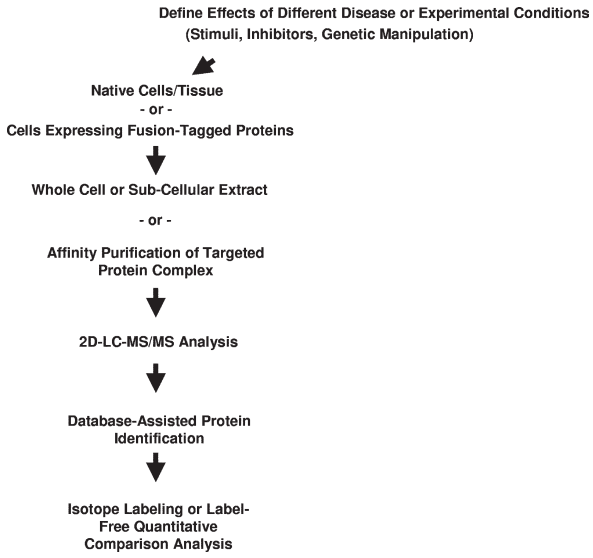

Figure 3. Defining regulatory events in targeted protein samples with 2D-LC-MS/MS analysis. The composition of whole-tissue/cell-extract or purified protein complexes is characterized using the 2D-LC-MS/MS approach described in Figure 1. Targeted protein complexes are purified from cell or tissue extracts using the different approaches described in Figure 2. Quantitative differences between experimental/disease conditions and controls are determined using isotope labeling (metabolic or chemical tags) or label-free approaches.

COMPUTATIONAL INTERPRETATION OF 2D-LC-MS/MS DATA

The earlier sections focused on how different types of biological samples are prepared for 2D-LC-MS/MS analysis. This section discusses the different informatic platforms that are required for interpreting the data acquired from 2D-LC-MS/MS experiments. Informatics is required to convert the mass spectral information into the identity of proteins in the biological sample and comparison of experimental/disease conditions versus controls. Informatics is used to map the mass spectral data to a peptide sequence that can be compiled into the identity of the proteins in the sample and for systematic comparison of experimental/disease conditions versus controls. A number of efforts are underway to develop and optimize informatics, starting with the software of the mass spectrometry instruments, analyzing the acquired spectra, and comparing the results of multiple experiments to generate models of protein function, protein structure, and disease predictions. For protein identification, the amino acid sequences from acquired MS/MS spectra are inferred by comparing the theoretical properties of peptides derived from translated genome sequences to the experimental mass spectrometry data (Fig. 1). Multiple computer algorithms have been developed that process the unedited MS/MS data into lists of peptide sequences that statistically correlate with the raw mass spectrometry data. The most commonly used algorithms are SEQUEST (Thermo Fisher Scientific, Waltham, MA), MASCOT (Matrix Science, Boston, MA), and X!Tandem (www.thegpm.org/tandem).[28-32] These search algorithms mathematically compare the experimental fragmentation ion (MS/MS) data with the predicted fragmentation ions of the peptide sequences to generate scoring metrics describing the fit of the experimental and theoretical data. By using different statistical approaches, the algorithm incorporates the scoring metrics to allow for ranking of the peptide sequences with the experimental MS/MS spectrum.

Most of the search algorithms attempt to re-assemble the identified peptide sequences into a list of proteins. Typically, the output of peptide sequences and scoring metrics is filtered, sorted, and organized into a list of the proteins that describe the initial protein composition.[2,17,31,33-36] Typically, a threshold value is set on scoring output from the protein database search results. Only peptides with scores above a user-defined threshold are allowed to pass through a filter into the final list of protein identifications. A biologist or clinician analyzing the 2D-LC-MS/MS results then must sift through these large numbers of protein identifications to evaluate their validity and biological significance. Needless to say, this is a very time-consuming process that tends to be very subjective. The final outcome typically is biased by the investigator's interpretation of the results. In the past few years, multiple statistical approaches, including linear discriminate analysis and machine learning algorithms, have been proposed for validating protein identifications from MS/MS results.[8,10,16,37-39] Comparative analysis across multiple experimental and control experiments is also essential for developing disease and biological models. Manual comparison of results containing hundreds to thousands of proteins is a daunting task. Thus, software applications have been developed for automated comparison of data from multiple 2D-LC-MS/MS experiments.[9,17]

QUANTITATIVE ANALYSIS OF PROTEINS FROM 2D-LC-MS/MS EXPERIMENTS

The effective assessment of a protein's contribution to a given cellular event or disease process requires that the relative abundance of proteins be determined accurately for quantitative comparison analysis of experimental/disease samples versus appropriate controls. Although 2D-LC-MS/MS identification of proteins from complex biological samples is becoming routine, the quantitative measurement of the components in these complex mixtures is more difficult to achieve. Quantitative mass spectrometry–based proteomics traditionally has relied on stable isotope labeling. Isotope labels are incorporated into proteins such that 2 or more sample states being compared are tagged with different mass variants of the label. By mixing the samples before mass spectrometry analysis, an assessment of the different mass ratio enables a pair-wise relative comparison of protein abundance levels. Sample mixing also eliminates the issue of experimental variability

that is associated with sample preparation and 2D-LC-MS/MS analysis. Metabolic labeling of cells before harvesting proteins and postisolation chemical labeling of digested proteins are 2 stable-isotope approaches that have been developed and refined for quantitative 2D-LC-MS/MS analysis.[11,40,41]

Metabolic labeling offers an elegant method for in situ quantitative analysis of proteins and protein networks in cultured cells. The elegance of this approach is that it allows comparative assessment of protein abundance changes in response to different disease-related or physiologic conditions. A proven metabolic labeling strategy involves introduction of stable isotope–labeled amino acids in cell culture, where proteins are tagged biosynthetically with stable isotopes during protein synthesis.[11] A number of schemes have been developed to introduce stable isotopes of hydrogen, carbon, nitrogen, and oxygen into a variety of different essential amino acids including arginine, tryptophan, cysteine, aspartic/glutamic acid, and lysine. Lysine or arginine often is the favorite choice for 2D-LC-MS/MS experiments because lysine- or arginine-terminated tryptic peptides presumably will contain a single stable isotope label. These reagents also are available commercially as different mass variants, allowing quantitative comparison of different cellular states in a single experiment. For example, examination of changes in targeted protein complex composition in response to activation or inhibition of a particular cytokine signaling pathway.

One of the first chemical labeling techniques used reagents called isotope-coded affinity tags (ICAT).[12] ICAT reagents contain a reactive group for conjugation to cysteine residues and a biotin moiety for selective isolation of tagged peptides using streptavidin beads. The original ICAT reagents consisted of a heavy form that contained 8 deuterium atoms, and a light form that contained no deuterium. A later version of the ICAT reagent used ^{13}C as the heavy isotope.[41] One of the drawbacks or concerns with using cysteine-targeted ICAT reagents is that peptides that do not contain cysteine residues will be missed in the analysis. To address this limitation, reagents that label at the N-terminus of all peptides have been developed. One of these reagents, termed *isobaric tags for relative and absolute quantitation* (iTRAQ), has become the method of choice for chemical labeling.[40] In addition to being amine-reactive, isobaric tags for relative and absolute quantitation reagents use fragmentation-tags of 4 different masses. This is advantageous because it enables up to 4 different sample conditions to be compared in one experiment. For instance, 4 different progression stages of a particular type of nephropathy can be monitored in a single experiment.

Because of the complexity of sample processing and the substantial costs associated with using stable-isotope reagents, approaches for quantifying proteins in complex biological samples using 2D-LC-MS/MS without the use of labels have been pursued.[3,13,14,20,39,42-44] Experiments showing a linearity of MS ion intensity or the number of spectral counts with peptide or protein concentration provides the framework for effective label-free quantitative analysis of 2D-LC-MS/MS experiments.[15,39,44] Two groups recently used these methods for enrichment-based assignment of proteins to specific subcellular compartments.[24,25] The utility of these methods were supported by Western blot and in situ imaging analysis and by consistency with benchmark localization datasets. Our group also has proposed a label-free method for estimating protein abundance from 2D-LC-MS/MS experiments using the spectral count principle. This method estimates the relative abundance of each identified protein by normalizing the number of spectral counts matching to the protein by its predicted molecular weight. We call this value a *protein abundance factor*. These data interpretation are facilitated using a web-based informatic system named BIGCAT (http://plato.kdp.louisville.edu/BIGCAT/index.php).[9] This label-free approach has been highly successful in the development of statistical models based on large sets of 2D-LC-MS/MS experimental data.[1,3,20,45]

APPLICATION OF 2D-LC-MS/MS ANALYSIS TO DEFINING REGULATORY EVENTS IN KIDNEY BIOLOGY AND DISEASE

The previous sections described some general background information pertaining to the use of 2D-LC-MS/MS analysis in defining regulatory

protein networks. Here we discuss some specific kidney-related applications of these methodologies.

Identification of Regulatory Signaling Events in Cultured Renal Cells

A key feature of this technology is that it allows an unbiased assessment of interactions within pertinent signaling networks and changes that are mediated by disease-related stimuli or pharmacologic inhibitors. Renal fibrosis is the final common manifestation of a wide variety of chronic kidney diseases (CKDs). This process is characterized by both glomerulosclerosis and tubulointerstitial fibrosis, which results from improper regulation of the body's wound-healing system. Thus, identifying protein–protein interactions that are mediated by proinflammatory cytokines (ie, tumor necrosis factor [TNF-α]) or other related stimuli in glomerular and tubular cells will help define mechanisms regulating this central disease process. Bouwmeester et al[18] used the TAP-2D-LC-MS/MS approach for extensive characterization of the TNF-α/nuclear factor κB (NF-κB) signaling networks. TAP was used to purify 32 previously characterized components of the TNF-α/NF-κB transduction pathway in HEK293 cells. Comparing replicate experimental TAPs with non–TNF-α signaling–related TAP experiments, they identified 131 high-confidence interactors, of which 80 were previously unknown interactors. Of these 80 candidates, 28 were selected for functional validation studies using a combination of small interfering (si)RNA knock-down and NF-κB transcription activity analysis. Ten of these candidates reproducibly regulated NF-κB activity. These findings suggest that the TAP approach offers a substantial level of identification sensitivity and functional specificity. Although this report shows the utility of the approach for characterizing targeted signaling complexes, the challenge has been effective quantitative assessment and comparison such that modulations mediated by altered cellular conditions can be determined. A recent study described a unique quantitative strategy for comparative analysis of LC-MS/MS results.[23] The previously described ion intensity-based label-free method was used to assess the effects of pharmacologic phosphoinositide-3 kinase (PI3K) inhibition (Ly294002) on protein interactions with the forkhead transcription factor Fox03A in HEK293 cells. This is a well-characterized network that was selected to show the validity and performance of the technology. Nonetheless, this signaling cascade is a key regulator of transforming growth factor (TGF-β) expression, a prominent mediator of fibrosis in a number of types of CKD.[46-48] The primary challenge of this type of analysis is differentiating specific changes apart from those resulting from normal experimental variably. To account for this, samples from PI3K inhibition and controls were mixed at increasing ratio increments (LY/control: 0%, 20%, 80%, and 100%) before LC-MS/MS analysis. Thus, proteins with abundance changes that were commensurate with this ratio were interpreted to be PI3K-mediated changes. These PI3K-altered interactions were validated using immunofluorescence colocalization experiments. The utility of this approach also was indicated by identification of previously characterized PI3K-mediated events. Although these studies were performed with HEK293 cells and 2 distinct protein complexes, these methods should be applicable to any transfectable renal cell line and other signaling complexes of interest.

Assessment of Protein Expression Profiles in Selective Cellular Components of the Kidney

Another important application of LC-MS/MS analysis is the identification of selectively enriched subsets of cellular proteins from animal tissue (reviewed by Andersen and Mann[49]). These studies provide valuable insight into in vivo protein function and regulation. One such recent report used centrifuge-based isolation coupled with 2D-LC-MS/MS to characterize the protein composition of 4 different subcellular fractions (cytosol, membrane, mitochondria, and nuclei) from a variety of mouse tissues including kidney.[25] Approximately 700 proteins from each kidney fraction were identified. Cellular localization specificity was indicated by protein abundance or enrichment as determined by a spectral counting–based, label-free quantitative approach described in a previous section. The validity of these assign-

ments was indicated further by concordance with benchmark datasets. A few of the uncharacterized assignments also were validated experimentally with immunofluorescence imaging. An annotated list of the identified proteins can be viewed at http://tap.med.utoronto.ca/~mts. Another study focused on the enrichment and identification of integral and nonintegral membrane proteins of rat renal collecting duct cells.[50] By covalently labeling specific medullary preparations with biotin the investigators were able to affinity-purify and differentially identify proteins present in apical versus basolateral collecting duct membranes. The purified proteins were first separated by sodium dodecyl sulfate–polyacrylamide gel electrophoresis, and digested gel sections were applied to 1-dimensional LC-coupled MS/MS analysis. This approach is used commonly as an alternative to a 2-dimensional LC-coupled MS/MS approach. However, the 2D-LC approach eliminates time-consuming, gel-related sample preparation and, although not compared extensively, presumably improves the overall sensitivity of protein identification. Seventeen integral and 44 nonintegral apical and 23 integral and 134 nonintegral basolateral membrane proteins were identified in this study. The localization of a few of the candidate proteins were confirmed using immunofluorescence confocal microscopy. This list of proteins was compiled with other previously identified medullary collecting duct membrane proteins and can be viewed at dir.nhibi.nih.gov/papers/IKem/imp/index.htm. These extensive protein localization maps aid in our understanding of normal kidney biology and the methodologies establish the framework for identifying proteins that are regulated in animal models of different kidney diseases.

Assessment of Protein Expression Patterns in Kidney Disease and Relevant Cellular Models

Identification of protein expression changes that are mediated by kidney disease–related stimuli in relevant in vitro cellular models greatly improves our understanding of the disease process and aids in the discovery of novel therapeutic targets. Progressive renal failure in CKD is typified by an accumulation of activated fibroblasts (myofibroblasts) and subsequent myofibroblast-mediated deposition of excessive extracellular matrix in the tubulointerstitium.[51,52] The exact origin of these myofibroblasts is not understood clearly, but there is emerging evidence that suggests tubular epithelial cell transdifferentiation into myofibroblasts (EMT) is a major contributor.[51,53-56] EMT is activated in animal models of renal fibrosis and in cell cultures by several profibrotic factors including albumin, angiotensin II, fibroblast growth factor, and TGF-β.[47,53,55-59] Of these, TGF-β plays a central role in the renal fibrosis associated with a number of CKDs.[48] Much of our understanding of the molecular mechanism in EMT comes from studies defining mediators of EMT in tumor metastasis.[59] One such study used matrix-assisted laser desorption ionization–time-of-flight–assisted 2D-LC-MS/MS coupled with the iTRAQ stable isotope labeling approach to define proteins that were expressed differentially during TGF-β–induced EMT in human lung tumor cells.[38] A total of 51 differentially expressed proteins (29 up-regulated and 22 down-regulated) were identified. A majority of the proteins up-regulated by TGF-β activity regulate cell migration, adhesion, and invasion, suggesting an invasive phenotype and indicating the utility of this approach. However, most of these proteins typically are expressed at relatively high abundance and although phenotypic EMT changes were detected by immunoblotting, none of these or other typical EMT markers were identified. This suggests sensitivity limitation of this approach. Arguably, electrospray-based linear ion traps are the most sensitive of the currently available mass spectrometers. However, the iTRAQ reagents were designed exclusively for use with Applied Biosystems time-of-flight–based mass spectrometers. Although attempts have been made to develop strategies to allow use of iTRAQ with ion trap instruments, these methods have not been applied routinely.

Identification of Novel Regulators of Diabetic Nephropathy

Our laboratory is using ion trap-assisted 2D-LC-MS/MS and spectral counting-based label-free

methods to identify global protein expression changes in whole renal tubular extracts from 6-month-old OVE26 type 1 diabetic and control mice. This diabetic mouse model and this time point was selected because at this age these animals display features of late-stage human diabetic nephropathy, including albuminuria and TIF.[60] Overall, we have identified over 14,000 different proteins with just 10 μg of tubular protein from replicate diabetic and control MS experiments. This has allowed us to begin mapping the renal tubular proteome (Cummins et al, unpublished data). Thus far, we have identified 477 significantly differentially expressed ($P < .05$) proteins in the diabetic tubular cells as compared to controls. Using Ingenuity Pathways analysis (www.ingenuity.com), histograms were generated to illustrate functional information for this subset of differentially regulated proteins (Fig. 4). Interestingly, there appears to be a functional trend in that many of these diabetes-regulated proteins are involved in glucose metabolism and energetics. It is biologically plausible that in the high-glucose environment of diabetes that these types of molecular processes would be regulated in this manner. We also observed differences for proteins with substantial biological relevance, but

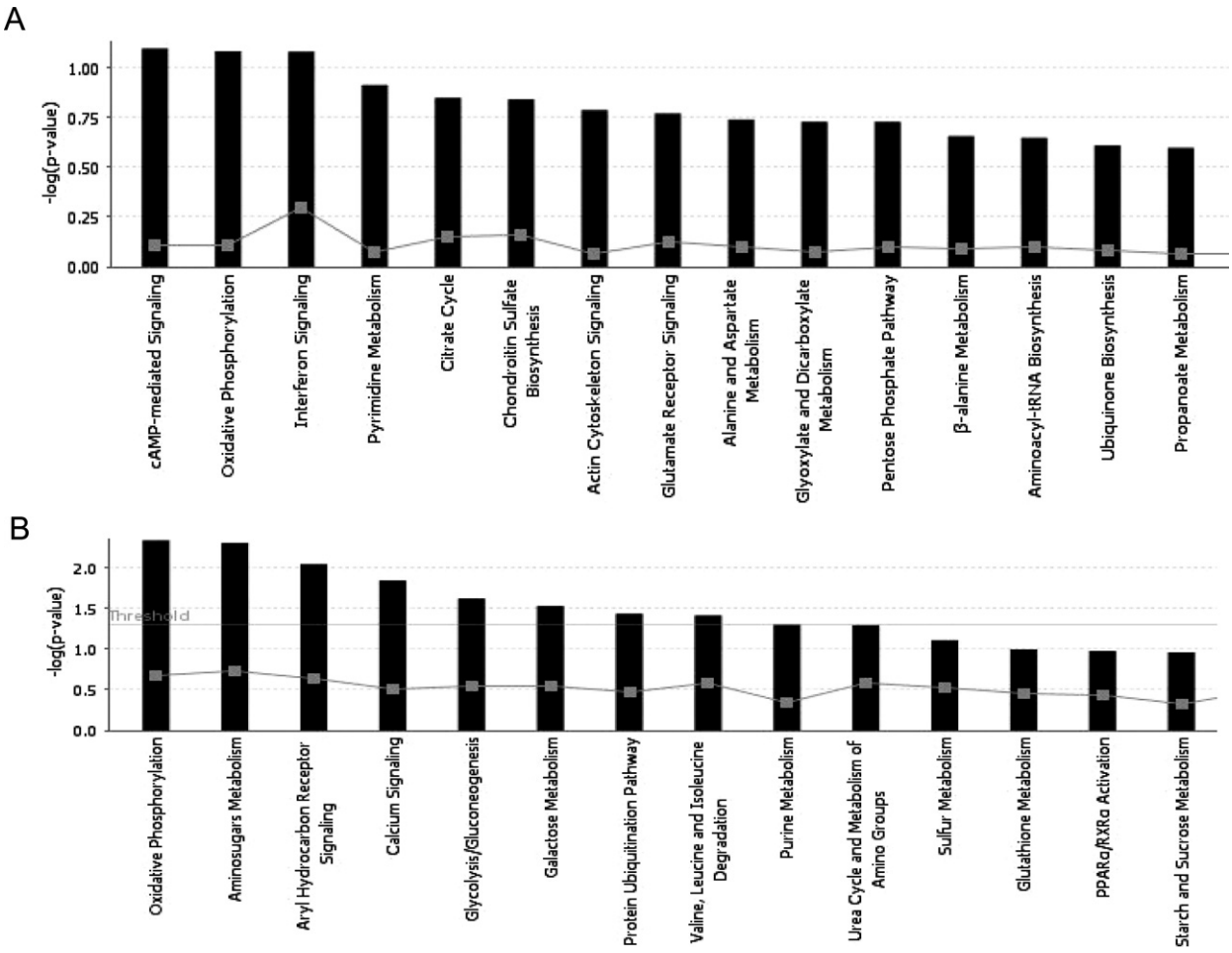

Figure 4. Histogram showing functional distribution of a subset of differentially expressed proteins in OVE26 diabetic mouse renal tubular cells. Replicate 2D-LC-MS/MS analyses were performed on renal tubular extracts from OVE26 diabetic and control mice. Protein abundances were determined using a previously described spectral counting-based approach.[3] Data filtering and quantitative comparative analysis was performed using an in-house web-based program.[9] The histogram shows groups of proteins that were significantly ($P < .05$) differentially expressed proteins in diabetic versus control samples. Each bar represents the functional group that maps back to the dataset and how well the data match that functional group with a significance threshold of $P < .05$. The top 14 functional groups are listed. (A) Functional categories for proteins increased and (B) decreased in diabetic extracts.

these had *P* values greater than .05. Many of these regulated proteins are EMT and fibrosis related proteins (Powell et al, unpublished data). Another diabetes-induced protein of particular interest was albumin. This may be because these diabetic animals display profound albuminuria starting at about 3 months of age.[60] In concordance with our MS results, immunohistochemistry showed enhanced albumin expression in the cortical tubules of diabetic mice (data not shown in lieu of Cummins et al, unpublished data). We propose that these biologically relevant findings indicate the utility of this method for unbiased discovery of novel regulatory events in kidney disease and that selective analysis of different components of the nephron will help define site-specific regulation.

REFERENCES

1. Fleischer TC, Weaver CM, McAfee KJ, et al. Systematic identification and functional screens of uncharacterized proteins associated with eukaryotic ribosomal complexes. Genes Dev. 2006;20:1294-307.
2. Link AJ, Eng J, Schieltz DM, et al. Direct analysis of protein complexes using mass spectrometry. Nat Biotechnol. 1999;17:676-82.
3. Powell DW, Weaver CM, Jennings JL, et al. Cluster analysis of mass spectrometry data reveals a novel component of SAGA. Mol Cell Biol. 2004;24:7249-59.
4. Powell DW, Merchant ML, Link AJ. Discovery of regulatory molecular events and biomarkers using 2D capillary chromatography and mass spectrometry. Expert Rev Proteomics. 2006;3:63-74.
5. Lominadze G, Powell DW, Luerman GC, et al. Proteomic analysis of human neutrophil granules. Mol Cell Proteomics. 2005;4:1503-21.
6. Yates JR 3rd, McCormack AL, Eng J. Mining genomes with MS. Anal Chem. 1996;68:534A-40A.
7. Craig R, Cortens JC, Fenyo D, et al. Using annotated peptide mass spectrum libraries for protein identification. J Proteome Res. 2006;5:1843-9.
8. Listgarten J, Emili A. Statistical and computational methods for comparative proteomic profiling using liquid chromatography-tandem mass spectrometry. Mol Cell Proteomics. 2005;4:419-34.
9. McAfee KJ, Duncan DT, Assink M, et al. Analyzing proteomes and protein function using graphical comparative analysis of tandem mass spectrometry results. Mol Cell Proteomics. 2006;5:1497-513.
10. Nesvizhskii AI, Keller A, Kolker E, et al. A statistical model for identifying proteins by tandem mass spectrometry. Anal Chem. 2003;75:4646-58.
11. Ong SE, Blagoev B, Kratchmarova I, et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics. 2002;1:376-86.
12. Gygi SP, Rist B, Gerber SA, et al. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol. 1999;17:994-9.
13. Wiener MC, Sachs JR, Deyanova EG, et al. Differential mass spectrometry: a label-free LC-MS method for finding significant differences in complex peptide and protein mixtures. Anal Chem. 2004;76:6085-96.
14. Old WM, Meyer-Arendt K, Aveline-Wolf L, et al. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. Mol Cell Proteomics. 2005;4:1487-502.
15. Wienkoop S, Larrainzar E, Niemann M, et al. Stable isotope-free quantitative shotgun proteomics combined with sample pattern recognition for rapid diagnostics. J Sep Sci. 2006;29:2793-801.
16. Radulovic D, Jelveh S, Ryu S, et al. Informatics platform for global proteomic profiling and biomarker discovery using liquid chromatography-tandem mass spectrometry. Mol Cell Proteomics. 2004;3:984-97.
17. Rauch A, Bellew M, Eng J, et al. Computational Proteomics Analysis System (CPAS): an extensible, open-source analytic system for evaluating and publishing proteomic data and high throughput biological experiments. J Proteome Res. 2006;5:112-21.
18. Bouwmeester T, Bauch A, Ruffner H, et al. A physical and functional map of the human TNF-alpha/NF-kappa B signal transduction pathway. Nat Cell Biol. 2004;6:97-105.
19. Gavin AC, Bosche M, Krause R, et al. Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature. 2002;415:141-7.
20. Sanders SL, Jennings J, Canutescu A, et al. Proteomics of the eukaryotic transcription machinery: identification of proteins associated with components of yeast TFIID by multidimensional mass spectrometry. Mol Cell Biol. 2002;22:4723-38.
21. Collins SR, Miller KM, Maas NL, et al. Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. Nature. 2007;446:806-10.
22. Butland G, Peregrin-Alvarez JM, Li J, et al. Interaction network containing conserved and essential protein complexes in Escherichia coli. Nature. 2005;433:531-7.
23. Rinner O, Mueller LN, Hubalek M, et al. An integrated mass spectrometric and computational framework for the analysis of protein interaction networks. Nat Biotechnol. 2007;25:345-52.
24. Foster LJ, de Hoog CL, Zhang Y, et al. A mammalian organelle map by protein correlation profiling. Cell. 2006;125:187-99.
25. Kislinger T, Cox B, Kannan A, et al. Global survey of organ and organelle protein expression in mouse: combined proteomic and transcriptomic profiling. Cell. 2006;125:173-86.
26. Taal MW, Omer SA, Nadim MK, et al. Cellular and molecular mediators in common pathway mechanisms of chronic renal disease progression. Curr Opin Nephrol Hypertens. 2000;9:323-31.
27. Lee KK, Florens L, Swanson SK, et al. The deubiqui-

tylation activity of Ubp8 is dependent upon Sgf11 and its association with the SAGA complex. Mol Cell Biol. 2005;25:1173-82.
28. Clauser KR, Hall SC, Smith DM, et al. Rapid mass spectrometric peptide sequencing and mass matching for characterization of human melanoma proteins isolated by two-dimensional PAGE. Proc Natl Acad Sci U S A. 1995;92:5072-6.
29. Craig R, Beavis RC. TANDEM: matching proteins with tandem mass spectra. Bioinformatics. 2004; 20:1466-7.
30. Yates JR 3rd, Eng JK, McCormack AL, et al. Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. Anal Chem. 1995;67:1426-36.
31. Yates JR 3rd, Eng JK, McCormack AL. Mining genomes: correlating tandem mass spectra of modified and unmodified peptides to sequences in nucleotide databases. Anal Chem. 1995;67:3202-10.
32. Perkins DN, Pappin DJ, Creasy DM, et al. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis. 1999;20:3551-67.
33. Han DK, Eng J, Zhou H, et al. Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. Nat Biotechnol. 2001;19:946-51.
34. Tabb DL, McDonald WH, Yates JR 3rd. DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. J Proteome Res. 2002;1:21-6.
35. Elias JE, Gygi SP. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat Methods. 2007;4:207-14.
36. Chen Y, Kwon SW, Kim SC, et al. Integrated approach for manual evaluation of peptides identified by searching protein sequence databases with tandem mass spectra. J Proteome Res. 2005;4:998-1005.
37. Gatlin CL, Eng JK, Cross ST, et al. Automated identification of amino acid sequence variations in proteins by HPLC/microspray tandem mass spectrometry. Anal Chem. 2000;72:757-63.
38. Keshamouni VG, Michailidis G, Grasso CS, et al. Differential protein expression profiling by iTRAQ-2DLC-MS/MS of lung cancer cells undergoing epithelial-mesenchymal transition reveals a migratory/invasive phenotype. J Proteome Res. 2006;5:1143-54.
39. Liu H, Sadygov RG, Yates JR 3rd. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Anal Chem. 2004;76:4193-201.
40. Ross PL, Huang YN, Marchese JN, et al. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Mol Cell Proteomics. 2004;3:1154-69.
41. Yi EC, Li XJ, Cooke K, et al. Increased quantitative proteome coverage with (13)C/(12)C-based, acid-cleavable isotope-coded affinity tag reagent and modified data acquisition scheme. Proteomics. 2005;5:380-7.
42. Gao J, Friedrichs MS, Dongre AR, et al. Guidelines for the routine application of the peptide hits technique. J Am Soc Mass Spectrom. 205;16:1231-8.
43. Gao J, Opiteck GJ, Friedrichs MS, et al. Changes in the protein expression of yeast as a function of carbon source. J Proteome Res. 2003;2:643-9.
44. Wang W, Zhou H, Lin H, et al. Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. Anal Chem. 2003;75:4818-26.
45. Hughes AL, Powell DW, Bard M, et al. Dap1/PGRMC1 binds and regulates cytochrome P450 enzymes. Cell Metab. 2007;5:143-9.
46. Samatar AA, Wang L, Mirza A, et al. Transforming growth factor-beta 2 is a transcriptional target for Akt/protein kinase B via forkhead transcription factor. J Biol Chem. 2002;277:28118-26.
47. Perlman A, Lawsin LM, Kolachana P, et al. Angiotensin II regulation of TGF-beta in murine mesangial cells involves both PI3 kinase and MAP kinase. Ann Clin Lab Sci. 2004;34:277-86.
48. Bottinger EP, Bitzer M. TGF-beta signaling in renal disease. J Am Soc Nephrol. 2002;13:2600-10.
49. Andersen JS, Mann M. Organellar proteomics: turning inventories into insights. EMBO Rep. 2006;7:874-9.
50. Yu MJ, Pisitkun T, Wang G, et al. LC-MS/MS analysis of apical and basolateral plasma membranes of rat renal collecting duct cells. Mol Cell Proteomics. 2006;5:2131-45.
51. Gilbert RE, Cooper ME. The tubulointerstitium in progressive diabetic kidney disease: more than an aftermath of glomerular injury? Kidney Int. 1999;56:1627-37.
52. Liu Y. Renal fibrosis: new insights into the pathogenesis and therapeutics. Kidney Int. 2006;69:213-7.
53. Liu Y. Epithelial to mesenchymal transition in renal fibrogenesis: pathologic significance, molecular mechanism, and therapeutic intervention. J Am Soc Nephrol. 2004;15:1-12.
54. Rastaldi MP, Ferrario F, Giardino L, et al. Epithelial-mesenchymal transition of tubular epithelial cells in human renal biopsies. Kidney Int. 2002;62:137-46.
55. Simonson MS. Phenotypic transitions and fibrosis in diabetic nephropathy. Kidney Int. 2007;71:846-54.
56. Yang J, Liu Y. Dissection of key events in tubular epithelial to myofibroblast transition and its implications in renal interstitial fibrosis. Am J Pathol. 2001; 159:1465-75.
57. Kalluri R, Neilson EG. Epithelial-mesenchymal transition and its implications for fibrosis. J Clin Invest. 2003;112:1776-84.
58. Lan HY. Tubular epithelial-myofibroblast transdifferentiation mechanisms in proximal tubule cells. Curr Opin Nephrol Hypertens. 2003;12:25-9.
59. Zavadil J, Bottinger EP. TGF-beta and epithelial-to-mesenchymal transitions. Oncogene. 2005;24:5764-74.
60. Zheng S, Noonan WT, Metreveli NS, et al. Development of late-stage diabetic nephropathy in OVE26 diabetic mice. Diabetes. 2004;53:3248-57.