

Proteomics in 2005/2006: Developments, Applications and Challenges

Jeffrey C. Smith, Jean-Philippe Lambert, Fred Elisma, and Daniel Figeys*

Ottawa Institute of Systems Biology and Department of Biochemistry, Microbiology and Immunology, Faculty of Medicine, University of Ottawa, 451 Smyth Road, Ottawa, Ontario, Canada K1H 8M5

Review Contents

Technology	4325
Microfluidic Tools	4325
Microarrays	4326
Mass Spectrometry	4326
Protein Quantitation	4327
Posttranslational Modifications	4328
Glycosylation	4328
Lipidation	4328
Acetylation and Methylation	4328
Ubiquitination and SUMOylation	4328
Phosphorylation	4329
Identification of Protein–Protein Interactions	4329
Mammalian Cells	4330
<i>Saccharomyces cerevisiae</i>	4330
Viruses	4331
Bacteria	4331
Identification of Protein–Macromolecule Interactions	4331
Chemical Proteomics	4332
Systems Biology	4333
Protein Biomarkers	4334
Collection and Handling of Serum and Plasma	4334
Protein Depletion from Plasma and Serum	4334
Solubility-Based Approaches	4334
Antibody-Based Approaches	4335
Analytical Techniques	4336
Lab-on-a-Chip	4336
Free-Flow Electrophoresis	4336
Application of Protein Arrays To Study Plasma and Serum	4336
Data Handling	4337
Bioinformatics for Proteomics	4337
Open Source Projects	4337
Quantitative Software Development	4338
Data Analysis, Data Management, and Data Exchange Standards	4338
Interaction Networks, MS Identification, and Databases	4338
Concluding Remarks	4339
Literature Cited	4340

The field of proteomics has continually evolved as our global understanding of cellular and biochemical systems has matured and as new research tools have been developed. Our 2002–2004 review (1) discussed some of the research that impacted the development of the proteomics field throughout that time; many of the research areas that were included have remained strong throughout the past 2 years. In particular, the mapping of protein–

protein interactions has led to the near completion of the yeast interactome as well as forays toward the interactome of other species. Technological developments have also been strong, with higher emphases being placed on gel-free approaches and quantitation. The elucidation of proteomic biomarkers has continued but has also faced growing controversy surrounding some of the technical applications used in the discovery process including the overfitting of data. Many different bioinformatic tools have been developed to aid research in this field such as optimize the storage and accessibility of proteomic data or statistically ascertain the significance of protein identifications made from a single peptide match; a great deal of open source software has also emerged. Here we attempt to provide a critical overview of the major developments in the field of proteomics that occurred during the years 2005–2006. Clearly, it is impossible to cover all of the literature that was published over these 24 months; we have therefore focused on a subset that is representative of the field and detail some of the success stories as well as some of the challenges that are currently being faced.

TECHNOLOGY

Microfluidic Tools. The field of proteomics has benefited from a number of technological advances over the past 2 years. Improvements to large-scale analytical techniques such as mass spectrometry and protein arrays have been reported; however, innovation in sample manipulation, particularly on a minute scale, has also allowed researchers to broaden their research scope and potential (2). To study one or a subset of proteins, affinity tags have been, and remain, commonly used; Lichty et al. (3) conducted an investigation into the benefits and drawbacks of different protein tagging strategies. They examined the efficiencies and cost of eight tagging methods and found substantial differences, concluding that the Strep II tag has the best overall performance for its associated cost. Several other review articles have also recently appeared, examining the benefits and limitations of protein tagging strategies (4–6). Advances in microfluidic techniques over the past 2 years have further enabled the study of low-abundance proteins or protein solutions with small (e.g., several microliters) volumes (7–9). Miao et al. (10) developed a rotating circular dichroism microfluidic platform whereby synthetic fusion proteins were purified using a chitin resin. The fusion proteins contained one or more inteins (self-splicing protein elements) as well as a chitin binding domain thereby allowing their purification on the resin and release through activation of the intein(s). Notably, this system allowed in vivo expression and purification of cytotoxic proteins as the inteins could be engineered

* To whom correspondence should be addressed. E-mail: dfigeys@uottawa.ca.

to disrupt the tertiary structure of the protein; fusion proteins were expressed and the native cytotoxic protein was released after purification and intein cleavage. Wheeler et al. (11) reported an extension of their electrowetting-on-dielectric based technique where analytes were moved across a Teflon surface, dried, desalted and had matrix applied to be analyzed by matrix-assisted laser desorption/ionization (MALDI) MS. Their technique successfully removed high concentrations of contaminants (e.g., 8 M urea) and improved the MALDI MS spectra considerably by removing adduct peaks. Moreover, they reported a substantial improvement in spectral signal-to-noise ratios as well as the number of peptides observed, leading to higher scoring protein identifications through peptide mass fingerprinting. Finally, a microfluidic device known as the "proteomic reactor" was reported by Ethier et al. (12). The reactor consisted of a small piece of fused-silica tubing packed with strong cation-exchange resin. Proteins bound to the resin and were reduced, alkylated, and digested within a few hours. The method not only increased the efficiency of the digestions (showing a 10-fold sensitivity increase compared to other microfluidic methods) but also allowed digestions to occur in a small volume (~50 nL); protein identifications were presented from the online lysis and digestion of 300 cells. Very recently, the utility of the proteomic reactor was demonstrated in the digestion and identification of immunoprecipitated proteins (13); this study allowed the rapid and efficient digestion of a simplified sample of proteins leading to the successful identification of a number of ubiquitination sites. A 96-well multiplexed version of the proteomic reactor was also recently developed whereby samples could be processed in parallel (14). This study reported an enhanced number of protein identifications through a size exclusion chromatographic separation of a cellular lysate followed by the digestion of several micrograms of each fraction on the multiplexed reactor. Peptides from each well were analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS) leading to several thousand protein identities, all derived from several hundred micrograms of cell lysate.

Microarrays. Proteomic experiments using microarray-based strategies have been reported numerous times over the past 2 years for a variety of applications. Microarray technologies are extremely sensitive; their importance is implicated in the number of recent reviews on this subject (15–27). Nettikadan et al. (28) developed a miniaturized antibody microarray (termed an "ultramicroarray") with a reduced spot size ranging from 1 to 20 μm (representing 1/100–1/10 000 the size of a spot on a traditional microarray). The smaller functionalized surfaces still produced measurable and statistically significant signals however substantially decreased the volume requirements necessary to probe and detect the antigens. The ultramicroarray was used to detect prostate serum antigen (PSA) secreted from four cells over a 24-h period as well as PSA in the cellular lysate of six cells. Aptamer microarray technologies have also been reported a number of times over the past 2 years. Collett et al. reported methods to produce and use aptamer microarrays (29) and expanded the discussion the following year by describing how these microarrays may be used to target and quantitate multiple proteins simultaneously (30). They described a multiplexed aptamer array that used both RNA and DNA aptamers to simultaneously target up to four proteins at a time. Moreover, they presented the details

of a rigorous experimental optimization of aptamer immobilization conditions, surface coating, and protein binding as well as labeling strategies. Steinhauer et al. (31) developed several porous silicon-based solid supports for antibody microarrays as well as nitrocellulose-coated variants of them, showing that they displayed properties equal to, or better than, the solid supports currently used in the field. Finally, an impressive project was published at the end of 2005 by Ptacek et al. (32), where proteome chips were used to map a large fraction of the protein phosphorylation events in yeast. The group had previously developed microarray chips containing the majority of yeast proteins individually immobilized to separate spots; yeast kinases were overexpressed, purified, and incubated on the proteome chips in the presence of radioactive ATP. Following incubation, slides were assayed for radioactivity to indicate which proteins were substrates to the kinase in question. Using this method, they discovered over 4000 phosphorylation events involving 1325 different proteins, more than eight times the number of phosphoproteins known at that time (32). Their report also compared the phosphorylation behavior of several substrates when treated with related kinases and included an interaction map of the results. A limitation of their approach, however, was that it could not determine the exact site of phosphorylation on each protein.

Mass Spectrometry. MS has arguably had the largest impact on the rise of proteomics over the past several decades. The majority of substantial technological advances in this field have been well established for some time; recent developments over the past 2 years have mainly included minor improvements to instrument electronics as well as the addition/modification of software functionality. There have also been several publications over the past 2 years that have detailed the development of novel MS hardware and investigated the advantages that they pose to various proteomic applications. The orbitrap mass spectrometer was reported as a tool for proteomics research in 2005 by Hu et al. (33). Invented by Makarov in 1999, ions are trapped and orbit around a central spindlelike electrode where they induce an image current in outer electrodes; the frequency of the current is Fourier transformed into the time domain producing mass spectra (34). Moreover, this new technology is easily coupled with common linear ion traps to produce hybrid instruments capable of a wide variety of different experiments (34). For example, Macek et al. (35) used an orbitrap instrument to accurately measure the mass of intact standard proteins, the linear ion trap component to collisionally dissociate various charge states of the protein, followed by the orbitrap to determine the masses and charge states of the fragment ions; the utility of this approach in widespread proteomic experiments was discussed. The use of modified linear ion trap mass spectrometers that allow electron-transfer dissociation (ETD) experiments to be conducted has also been recently introduced (36) and discussed in the context of proteomic analysis by MS, particularly in the area of posttranslational modifications (37). It has been demonstrated in many cases that ETD MS/MS provides superior sequence coverage compared to traditional collision-induced dissociation MS/MS when analyzing phosphorylated or methylated peptides owing to the fact that the ion dissociation mechanisms involved in ETD do not invoke the loss of the labile PTM; the internal energy gained

in the ETD process translates into peptide backbone fragmentation rather than PTM cleavage (37).

PROTEIN QUANTITATION

There has been a push in the proteomics literature to identify proteins over the course of several conditions and elucidate their relative or absolute temporal dynamics. This has been facilitated, in part, by a host of strategies currently available to conduct quantitative proteomics as well as the realization that cellular systems may be better understood as the proteome dynamics in response to external or internal stimuli are determined. Liu et al. (38) recently developed a technique entitled metal element chelated tags where the bicyclic anhydride diethylenetriamine-*N,N,N',N',N''*-pentaacetic acid is covalently coupled to peptide primary amines and chelated to the rare earth metals Y and Tb. Tagged peptides are mixed and quantified based on the relative intensities of the Y and Tb tag pairs; this work was reported as a variation of the element-coded affinity tagging strategy developed in 2004 (39), however, was only demonstrated on protein standards, not on proteomic samples generated from a cellular lysate or immunoprecipitation. Hoffert et al. (40) utilized software to identify and quantify phosphopeptides of renal cells in response to vasopressin treatment in a label free manner. The software compared the chromatographic peak areas of phosphopeptides detected by MS to construct relative signal intensities and thus infer concentration differences without the need for chemical or metabolic labeling. This software, termed Quoil (*Quantification without isotopic labeling*), was simultaneously described in detail in a separate publication by Wang et al. (41) Similar quantitative software programs that determine relative peptide concentrations based on MS signal intensities in a label-free manner have been reported in the last 2 years by Johansson et al. (42) and Pan et al. (43) as well as Hattan and Parker (44); the latter also included an algorithm that considered LC retention time when grouping proteins from subsequent experiments. Hendrickson et al. (45) compared a software-based quantitative technique to a more standard metabolic labeling data set that contained proteome and transcriptome measurements from the same organism. They concluded that spectral counting showed an overall lower sensitivity defined in terms of detecting a 2-fold change in protein abundance compared to the metabolic methods.

There have been many reports over the past 2 years on the application of well-established quantitative proteomic strategies to investigate a variety of different biological systems. A complete description of this literature is well beyond the scope of this review and has been described previously (1, 46); however, a selection will be briefly outlined. The AQUA (absolute quantification) strategy has been recently used to study protein post-translational modifications including phosphorylation (47) and ubiquitination (48, 49); however, AQUA remains a fairly expensive technique that requires prior knowledge of the analyte in question. Isotopically labeled dimethyl labeling is a relatively inexpensive method of quantitation and has been used to monitor the dynamics of protein phosphorylation in rat uteri (50) as well as proteins involved in Alzheimer's disease (51). Dimethyl labeling has also been shown to increase the abundance of the a_1 ion, which may be used as a mass tag to increase the confidence in peptide identifications (52). Isotope-coded affinity tags

(ICAT) have been used to monitor the proteomic dynamics of growth factor compositional changes of lipid rafts in human smooth muscle cells (53) as well as hippocampal postsynaptic density-associated proteins following morphine administration (54). A newer version of the ICAT tag that is acid cleavable was introduced in 2006 by Flory et al. (55) and was demonstrated on the cell cycle of budding yeast. The isobaric tag for relative and absolute quantitation (iTRAQ) strategy (56, 57) has been used in a variety of interesting areas including neuronal development and dysfunction (58–60), phosphoproteomics (61), microsomal physiology (62), and anthrax biology (63). Wu et al. (64) recently conducted a comparative study between the iTRAQ, ICAT, and difference gel electrophoresis (DIGE) strategies, concluding that, although more susceptible to errors due to inaccurate precursor ion selection, iTRAQ was the most sensitive of the three followed by ICAT, which had a very similar sensitivity to DIGE despite being limited to cysteine-containing proteins. In general, the three techniques were found to be complimentary; however, the reproducibility of each technique was not reported and the variation that is observed through replicate analyses (65) could have accounted for some of the differences that were observed.

Stable isotope metabolic labeling methods have gained tremendous popularity over the past 2 years owing to an increased commercial availability of reagents, the fact that derivatization methods are not required following protein isolation, the inherent interexperimental reproducibility associated with these techniques (66), and the suitability of these approaches to a wide variety of in vitro experiments. Regular medium is easily replaced with metabolically labeled medium in culture dishes or flasks, allowing cellular growth, protein isolation, digestion, and analysis; MS analysis of peptides derived from differently labeled media contain amino acids of distinct mass and thereby allow quantitation. Stable isotope labeling by amino acids in cell culture (SILAC) (67, 68) is perhaps the most popular method of metabolic labeling currently used in the scientific community; detailing every report that utilized this technique over the past 2 years would be a review unto itself, however, a few interesting examples will be briefly highlighted. SILAC has been used to study proteomic properties of cancer cells including the dynamics of cancer progression as well as cancer cell death (69–74). It has provided insight into phosphoproteomic dynamics (75–77), protein turnover rates (78), the biochemistry of cells involved in the immune response (79, 80), proteins involved in DNA damage response (81), and plant proteins involved in abiotic stress response (82).

Finally, an innovative non MS-based quantitative strategy was recently reported by Drukier et al. (83) where blood proteins were quantified in an ELISA-type manner. Using their "ultrasensitive immunoassay", antibodies were coupled with antigens, which were then probed with a secondary antibody containing ^{125}I -streptavidin, rather than the traditional streptavidin/enzyme conjugate. The radioactive iodine-containing streptavidin decays by an electron capture mechanism whereby multiple photons/particles are emitted simultaneously creating a distinct radiation pattern and allowing background noise to be efficiently subtracted; this process is known as multiphoton detection (MPD). Using MPD, as little as 1000 atoms of ^{125}I may be detected, increasing the sensitivity of the technique 1000-fold over traditional immunoassay

methods. Examples detecting analytes with concentrations as low as 10 fg/mL (40 zmol/mL) were presented.

POSTTRANSLATIONAL MODIFICATIONS

The identification of protein posttranslational modifications (PTMs) represents a very challenging field that has received a great deal of attention over the past 2 years. Again, a full review of the work conducted in this area is beyond the scope of this review; alternatively, we have selected a few groundbreaking and innovative examples that detail novel approaches to study some of the most common PTMs; additionally, several interesting applications using previously established techniques will be presented.

Glycosylation. Protein glycosylation represents the most common PTM in mammalian cells and is challenging to study owing to the heterogeneity of different glycoforms that may be covalently attached to proteins (84–87). There have been a number of developments to improve the affinity capture and purification of glycopeptides prior to analysis by MS. Lectin affinity chromatography has been reported numerous times as an efficient method of capturing glycopeptides. Yang et al. (88) reported one of the first large-scale glycoproteomic experiments utilizing multilectin affinity chromatography, analyzing human serum and plasma and identifying approximately 150 different glycosylated proteins. In this approach, glycoproteins were isolated by agarose-bound lectins immobilized within the column and eluted following purification using a solution containing simple sugars. Several large-scale reports based on the same experimental principles have followed including a comparative analysis of sialylated glycoproteins from normal serum (89) as well as normal and pancreatic cancer serum (90). Okanda and Rassi (91) developed monolithic lectin affinity columns and used several of them in tandem to achieve superior separations of glycosylated proteins. Vosseller et al. (92) developed a weak lectin affinity chromatography method whereby glycopeptides were not permanently bound to the lectin resin but had retarded elution times owing to the affinity between the two phases allowing their enrichment and identification by MS. Glycopeptides were first isolated on the basis of hydrazide chemistry in 2003 (93), and several large-scale studies have been conducted using this technique in the past 2 years including a high-throughput quantitative analysis of serum proteins (94), a quantitative analysis of cisplatin-resistant ovarian cancer cells (95), and a comprehensive glycoproteomic analysis of human saliva (96). Finally, a novel method of glycoprotein capture was reported by Sprung et al. (97) as a further application of their “tagging via substrate” (TAS) strategy developed for isolating farnesylated proteins (98) whereby cells were cultured in the presence of an unnatural GlcNAc analogue. The GlcNAc analogue was metabolized by the cells, and glycosylated proteins were conjugated *in vivo* to a biotinylated capture reagent via the Staudinger ligation, allowing protein purification using a streptavidin affinity column. The TAS strategy was demonstrated on 4 different cell lines; however, only 41 glycosylated proteins were reported in total. Surprisingly, most glycoproteomic studies only identify dozens to hundreds of glycoproteins despite estimates that over half of a given proteome is glycosylated; clearly, superior methods of isolating and identifying this PTM need to be developed to allow comprehensive analyses in the future. Moreover, unambiguous

glycopeptide identification with respect to both the peptide and glycoform sequences is still not routine in high-throughput studies illustrating the need for further innovation in this field.

Lipidation. Lipoproteomics (lipid groups covalently attached to proteins) is a relatively small field compared to the study of other protein PTMs. This is likely due to a variety of experimental hurdles associated with the study of lipoproteins such as the high molecular weight of many lipid groups and that lipidation may significantly alter the chromatographic properties of proteins or peptides; in general, lipoproteins/peptides are not easily analyzed by MS and have largely been ignored. Lipoproteins play a variety of important roles within the cell, however, and deserve a great deal of attention if cellular systems are to be fully understood. Recent work in this field has been mainly limited to relatively simple methods compared to those used to study other PTMs. For example, Nally et al. (99) used MS to compare the measured masses of intact proteins isolated from outer membrane vesicles to their theoretical masses to estimate which lipid modifications were present. Hoffman and Kast (100) conducted MS/MS experiments on synthetic myristoylated, palmitoylated, and farnesylated peptides to elucidate fragment ions characteristic of the presence of each modification; the complexity of this field was further highlighted as relatively few marker ions were observed. A general strategy for detecting protein palmitoylation was also described whereby free sulfhydryl groups were blocked followed by the removal of palmitoyl groups and attachment of a radio-labeled or biotinylated reagent for detection or capture, respectively (101, 102).

Acetylation and Methylation. Very few novel methods of detecting protein methylation or acetylation have been reported in the past 2 years; most literature has focused on identifying this PTM using MS. MS-based strategies have been used to detect acetylation and methylation in human histones (103), N-terminal acetylation in archaea (104), methylation and dimethylation in human hair (105), methylation and acetylation variation between different cytogenetic leukaemia risk groups (106), and lysine methylation in HeLa cells as well as mouse liver mitochondria (107). A novel anti-methyllysine antibody was developed by Iwabata et al. (108) and used in conjunction with a previously developed anti-acetyllysine antibody on a variety of organ tissues in mice to determine the methylation and acetylation patterns throughout the organism.

Ubiquitination and SUMOylation. Ubiquitin and small ubiquitin-like modifier (SUMO) proteins are relatively small proteins that covalently attach to the side chain of lysine residues via an isopeptide bond; in so doing, they dramatically affect the behavior or fate of the protein to which they are attached. Their activity is critical in a variety of cellular processes, most notable perhaps, is the role that polyubiquitination plays in protein degradation (109, 110). A unique advantage in studying these PTMs is that they are proteins themselves, opening up a plethora of biochemical approaches with which to study them. In both ubiquitination and SUMOylation, tryptic digestion results in several amino acids remaining bound to the lysine residue at the site of modification, allowing a unique signature to be observed through MS/MS sequencing. There has been an increase in the number of studies on these PTMs in the past 2 years, the majority of which relying on different affinity capture strategies either by *in vivo* tagging

or by immunoprecipitation for capture of modified proteins. Ubiquitin molecules have been engineered to contain a tandem affinity histidine/biotin tag (111) or solely a histidine tag to be purified using metal affinity chromatography (112); both techniques may be conducted in denaturing conditions to reduce the probability of identifying nonubiquitinated proteins bound in a complex to another modified protein(s). Similarly, SUMO proteins have been modified *in vivo* to contain a polyhistidine tag for affinity purification (113); tandem affinity purification has been achieved on SUMOylated proteins using both histidine/FLAG tags (114) and histidine/S tags (115). In all cases, tagged proteins behave normally *in vivo*, binding to their target proteins, which are isolated as the ubiquitin or SUMO molecules are captured. Finally, immunopurification methods have been recently used to capture both monoubiquitinated and polyubiquitinated proteins from human cells. Vasilescu et al. (116) and Matsumoto et al. (117) used the FK2 antibody (which does not recognize free ubiquitin) to immunoprecipitate ubiquitinated proteins and identify them by MS. Interestingly, the latter study identified approximately five times as many proteins as the former, likely due to a number of factors including cell type differences (MCF-7 versus HEK293T, respectively) and the accuracies of the mass spectrometers used (QSTAR versus LCQ, respectively).

Phosphorylation. Protein phosphorylation is an important molecular switch that has received a great deal of attention in the scientific community for decades. Its prevalence in biological systems and role in disease states has led many researchers to attempt to identify which proteins are phosphorylated as well as the temporal dynamics of this PTM under many different conditions (86). A modest estimate of the number of reports in this area in the past 2 years alone is greater than 270; clearly only a subset of the most innovative phosphoproteomic strategies will be mentioned in the remainder of this section. Immobilized metal affinity chromatography (IMAC) was first used to isolate phosphoproteins over 20 years ago. The rise in popularity of phosphoproteomic studies in the past 2 years has led to the development of a number of improvement strategies in this area. For example, the entire IMAC process has been automated and coupled online to LC-MS/MS instrumentation (118), metal chelating resins have been replaced with cellulose powder (119), silica monolithic supports (120) or poly(glycidyl methacrylate/divinylbenzene) derivatized with iminodiacetic acid (121), phosphopeptide capture has been enhanced by the addition of 40% 1,1,1,3,3,3-hexafluoro-2-propanol (122), and methods to reuse IMAC columns have been developed. (123) Wolschin et al. (124) reported a technique similar to IMAC using metal oxide molecules for phosphopeptide capture. Al(OH)₃ was shown to successfully capture phosphopeptides from both standard phosphopeptide solutions and plant leaf protein extracts; unfortunately, the efficacy of this method was not directly compared to traditional IMAC methods. A number of strategies to detect phosphopeptides using different mass spectrometry scanning methods have also been developed in the past 2 years. Arrigoni et al. (125) replaced the phosphate group on phosphopeptides with an S-ethylpyridyl group while Amoresano et al. (126) dansyl-labeled phosphopeptides; both studies reported enhanced precursor ion(s) (*m/z* 106 and 170 or 234, respectively) as well as neutral loss transitions that could be used to trigger MS/MS fragmentation on the parent mass until satisfactory mass spectra

were obtained. Unwin et al. (127) developed a technique known as multiple reaction monitoring-initiated detection and sequencing (MIDAS) and used it to identify phosphopeptides associated with the cell cycle regulatory protein cyclin B from yeast. In the MIDAS strategy, multiple reaction monitoring (MRM) transitions are precalculated based on user input; for example, MRM losses of 98 Da were monitored for the known cyclin B phosphopeptide masses. MRM shows a dramatic sensitivity increase over other MS scanning methods as 100% of the instrumental duty cycle is used to monitor the transition, rather than only a fraction as is the case in parent ion, precursor ion, and neutral loss scanning. As an MRM transition is observed, the instrument switches into MS/MS mode to sequence and identify the phosphopeptide. The MIDAS strategy is valuable for samples of known and simple composition; however, it is less useful for global phosphoproteomic studies; monitoring hundreds to thousands of MRM transitions would divide the duty cycle between each transition, significantly diluting the advantage of the approach.

Other interesting developments in the field of phosphoproteomics have included the aforementioned work by Ptacek et al. (32), where 4000 phosphorylation events involving 1325 different proteins were detected in yeast using *in vitro* kinase assays on proteome chips (see above for details). Gevaert et al. (128) developed a strategy involving reversed-phase (RP) diagonal liquid chromatography where phosphopeptides were isolated with IMAC and separated into fractions by RPLC. Each fraction was then treated with a phosphatase and separated a second time by RPLC; the hydrophobicities of the dephosphorylated peptides increased as a result of the phosphate removal and showed retarded elution times relative to non-phosphopeptides in the fraction, facilitating their isolation and sequencing by MS. Gygi and co-workers have reported several statistical tools for phosphoproteomic studies in the past 2 years including a probability-based approach that determines the likelihood of correct phosphorylation site localization based on the presence and intensity of specific ions in the MS/MS spectra (129), as well as an iterative approach to determine phosphorylation motifs from large phosphoproteomic data sets (130).

IDENTIFICATION OF PROTEIN-PROTEIN INTERACTIONS

Through technology developments and methods standardization, the mapping of protein-protein interactions has matured over the past 2 years and the identification of protein interaction partners has become a standard assay in numerous laboratories. It is now a widely accepted view that the identification of protein-protein interactions can help define potential functions of an uncharacterized gene by linking it to proteins with known functions and expand our understanding of biological systems. Examples of small-scale experiments are abundant in the literature and are too vast to be covered in this review. In the last 2 years, numerous large-scale projects focusing on the identification of protein interaction partners in organism models such as *Saccharomyces cerevisiae*, *Escherichia coli*, and mammalian cells have been completed. These efforts have produced a wealth of information that can now be mined by researchers worldwide to help functionally annotate proteomes. The following section will focus on large-scale experiments and on reports describing significant improvements to the field.

Mammalian Cells. Wang et al. mapped the protein–protein interactions of Nanog, a transcription factor involved in stem cell proliferation and self-renewal, in mouse embryonic stem cells (131). Using a FLAG and streptavidin tandem affinity purification protocol, the authors were able to define Nanog interaction partners in stem cells. The authors validated their observations using coimmunoprecipitation and reciprocal affinity purification, enabling the creation of a protein interaction network enriched for proteins known to be involved in cell differentiation, development, and survival. The information gathered in this network was then used to investigate factors that cause mouse embryonic stem cells to differentiate into adult fibroblasts (132).

We recently mapped the protein–protein interactions of 338 proteins in human embryonic kidney cells by immunoprecipitation and high-throughput mass spectrometry (133). Briefly, the bait proteins were selected based on their association with a human disease or enzymatic functions, were engineered to contain a FLAG tag, and were transiently expressed in HEK293 cells. Bait proteins and their interaction partners were systematically immunoprecipitated using an anti-FLAG antibody, resolved on SDS-PAGE gels and visualized with colloidal Coomassie stain. In total, 1034 immunoprecipitation experiments were performed, resulting in the analysis of 16 321 gel bands by LC–MS/MS. A novel data-processing pipeline was implemented to handle the wealth of data produced in these experiments, which enabled the removal of low-quality data points as well as redundancy from the data set. The protein–protein interactions were filtered to remove nonspecific interaction partners (proteins present in more than 2.5% of the control experiments or interacting with more than 5% of the bait proteins under study). Finally, each protein–protein interaction was scored using six predictor variables enabling the computation of an interaction confidence score. In total, 6463 protein–protein interactions (2251 of which with high interaction confidence scores) were reported, constituting the largest protein–protein interaction study in human cells using mass spectrometry.

Barrios-Rodiles et al. (134) developed a novel assay to map protein–protein interactions termed luminescence-based mammalian interactome mapping or LUMIER. LUMIER uses the *Renilla* luciferase enzyme fused to a bait protein, which is then coexpressed with a FLAG tagged prey protein in mammalian cells. Following immunoprecipitation of the prey protein with anti-FLAG agarose beads, the interaction between the bait and prey proteins can be defined using a standard luciferase assay. The authors probed the transforming growth factor- β (TGF β) pathway with LUMIER to better comprehend the physiological effects of TGF β addition. In total, 518 FLAG tagged prey proteins were tested with 23 luciferase tagged bait proteins resulting in approximately 12 000 experiments. Of particular interest was the dynamic nature of numerous interactions that were either gained or lost following TGF β treatment.

Rual et al. (135) performed a systematic study of human protein–protein interactions within the human ORFeome v1.1, which contains approximately 8100 gateway-cloned open reading frames (ORFs) for approximately 7200 unique proteins, obtained using the yeast-two-hybrid (Y2H) method. The authors tested each possible pairwise interaction (7200 baits against 7200 preys) and reported ~2800 protein–protein interactions with a verification rate of ~78%. Stelzl et al. also performed a large-scale Y2H study

of human protein–protein interactions (136). They tested 4456 bait and 5632 prey proteins and observed 3186 interactions with a verification rate of 62% by membrane coimmunoprecipitation and 66% by pull-down experiments. Using more stringent criteria for the analysis of their interaction data, the authors reported 911 interactions with high confidence. To gain insight into the degeneration of cerebellar Purkinje cells, a frequent cause of ataxia in humans, Lim et al. mapped the protein–protein interactions of 54 proteins known to be involved in ataxia by Y2H (137). They screened the 54 bait proteins against the human ORFeome v1.1 (as per Rual et al. (135)) as well as human brain cDNA libraries and reported 770 protein–protein interactions. The authors then proceeded to validate a random fraction of their novel protein–protein interaction data by coimmunoprecipitation and successfully reproduced 62 out of 75 interactions (~83%).

A method termed quantitative immunoprecipitation combined with knockdown was recently developed by Selbach and Mann to reduce the false positive rate in the detection of protein–protein interactions (138). The methods relied on the SILAC labeling strategy and small interfering RNA to only identify interaction partners that specifically interacted with the bait proteins. The strategy involved two cell populations that were grown in light or heavy SILAC media; one sample was treated with siRNA against the bait protein of interest and immunoprecipitation was performed. The samples were then combined in a 1:1 ratio and analyzed by mass spectrometry. Interaction partners were identified through a decreased signal intensity in the siRNA sample when compared to the nontreated sample.

Saccharomyces cerevisiae. One of the major proteomic achievements in the last 2 years was the “complete” mapping of protein–protein interactions in *S. cerevisiae* by two independent groups (139, 140). Both groups systematically tagged most of the yeast proteins with a TAP tag that contained the *Staphylococcus aureus* protein A sequence and a calmodulin binding peptide separated by a cleavage site recognize by the TEV protease. Following a two-step immunoprecipitation procedure, the bait protein interaction partners were identified using common gel electrophoresis and mass spectrometry techniques. These efforts successfully identified the interaction partners for 1993 (140) and 2357 (139) bait proteins. These massive undertakings enabled many novel protein complexes to be identified, resulting in the assignment of cellular functions to previously uncharacterized proteins. One important difference between both groups was the bioinformatic methods used to analyze the raw protein–protein interactions data. Krogan et al. (139) used machine learning algorithms to reduce the noise in their protein–protein data that were trained using the MIPS curated protein interaction database (mips.gsf.de/). The authors then proceeded to use the Markov cluster (MCL) algorithm to identify protein complexes from their data (141). The MCL algorithm enabled the detection of 547 protein complexes, many of which were previously unknown. Gavin et al. (140) used a socioaffinity index to define the likelihood of an interaction between two proteins based on the probability of them being identified regardless of which one was used as the bait (spoke model) or copurifying when other proteins are tagged (matrix model). Using the computed socioaffinity indexes, the authors were able to make a matrix of all protein entries that could then be clustered to identify protein complexes. They subtracted

penalty values from the original matrix and repeated the clustering in an iterative manner to identify additional complexes (since proteins can be observed in more than one complex). A total of 491 protein complexes were identified using this approach, more than half of which were not previously known. Interestingly, a recent report has collapsed both data sets into one and reanalyzed it, resulting in the identification of 4456 novel protein–protein interactions (142).

While these colossal projects have attempted to study the yeast proteome on a global level, more focused studies can often elucidate additional novel and important information. One such study was designed to better understand the proteomics involved in translation regulation (143). Fleischer et al. adopted an integrative approach to purify yeast ribosomes using sucrose gradient sedimentation and validated their results by tandem affinity purifications. Furthermore, they studied the role of a fraction of the 77 novel proteins that were found to be associated with ribosomes (termed translation-machinery-associated proteins) and observed through deletion experiments that various members play a role in translation rate, fidelity, sensitivity to drugs, and polyribosome profiles.

Viruses. To better comprehend herpes virus infections, Uetz et al. mapped the protein–protein interactions in Kaposi sarcoma–associated herpesvirus (KSHV) and varicella-zoster virus (VZV) using Y2H (144). This enabled the identification of 123 and 173 interactions within the KSHV and VZV proteomes, respectively. It was observed that, in both viruses, the frequency of proteins that had few interaction partners was lower than what has been reported in either the yeast or the human interactomes to date. The authors proposed that this phenomenon was due to the “incompleteness” of the proteomes of the viruses and their dependency on host cell proteins for vital cellular functions. Mayer et al. studied the interactions between influenza virus ribonucleoproteins and mammalian cells (145). Specifically, they used purified influenza ribonucleoprotein complexes as bait and probed HEK 293 cell extracts for interaction partners, which were then identified by mass spectrometry. The authors also performed immunoprecipitation using a TAP tagged influenza A polymerase construct in HEK 293 cells to identify other interaction partners of influenza ribonucleoproteins. These experiments enabled the identification of 41 interaction partners for influenza ribonucleoproteins, some of which were validated by coimmunoprecipitation and colocalization experiments.

Bacteria. Bacteria have a long history of being used as model organisms in biological research; surprisingly however, only a few protein interaction studies have been conducted in *E. coli*. Butland et al. demonstrated that the yeast TAP tag system could be adapted to bacteria using the λ general recombination system (146). The authors attempted to tag 1000 open reading frames in the *E. coli* and succeeded for more than 800 proteins. Affinity purification resulted in the successful identification of 648 bait proteins as well as their interaction partners using mass spectrometry. Arifuzzaman et al. systematically tagged *E. coli* proteins with a hexahistidine-tag, overexpressed them for 2 h, and performed immunoprecipitation experiments using Ni²⁺-NTA columns (147). Interaction partners were identified for 2337 tagged proteins by mass spectrometry, enabling potentially conserved protein complexes to be studied. Furthermore, 798 uncharacter-

ized proteins were identified through their work, providing definitions for their cellular functions.

IDENTIFICATION OF PROTEIN–MACROMOLECULE INTERACTIONS

In recent years, the study of protein interactions has been extended to other macromolecules such as DNA and RNA in order to better define and understand biological systems. Of particular interest to numerous researchers are the interactions between protein complexes and condensed DNA or chromatin which are involved in fundamental cellular processes such as mitosis. For example, Riedel et al. mapped the interaction partners of the shugoshin (Sgo1) protein in *S. cerevisiae* (ScSgo1) and *Schizosaccharomyces pombe* (SpSgo1) using the TAP tag procedure (148). To improve the recovery of Sgo1-containing protein complexes on chromosomes, cells were synchronized by Pat1 inactivation and were harvested prior to meiotic division, which led to the identification of five proteins associated with Sgo1 in yeast. The authors proceeded to demonstrate that this complex was also observed in HeLa cells using immunoprecipitation with a mouse TAP tagged Sgo1. Denis et al. identified the interaction partners of Brd2, a protein with two bromodomains that are known to bind methylated histones, associated with chromatin (149). HA tagged Brd2 was immunoprecipitated using buffers containing ATP and analyzed by LC–MS/MS resulting in the identification of 28 proteins. These interaction partners were highly enriched for proteins related to chromatin structure, transcription factors, and cell-cycle regulators.

Histones are a group of proteins that physically interact with DNA and have a dramatic effect on chromatin structure and function. Core histones (histone H2A, H2B, H3, and H4) form an octameric nucleosome upon which DNA is wrapped, resulting in the compaction of DNA and restricting the ability of RNA polymerase to access it (150). Histone variants have been identified in numerous organisms for all core histones with the exception of histone H4; their functions are not completely understood (151). To better comprehend the role of a histone H3 variant in human cells, Foltz et al. performed tandem affinity purification of CENP-A using a tag containing protein A and S-protein sequences separated by a TEV cleavage site (152). Following a mild micrococcal nuclease treatment to reduce DNA size to approximately 150 base pairs, CENP-A containing nucleosomes were immunoprecipitated and their interaction partners were identified by mass spectrometry. Three novel centromere proteins were identified in this manner, two of which were used to perform reciprocal affinity purification using the same strategy. Further validation experiments were performed to ascertain the biological functions of the novel centromere components. Du et al. performed a similar proteomic study on the histone H2A variant, H2A.X, to identify its interaction partners when bound to DNA (153). H2A.X had been previously determined to be associated with DNA double-strand breaks; the experiments were designed to further explore the role of H2A.X in this process. They performed immunoprecipitation experiments of FLAG tagged H2A.X following a DNase I treatment and various exposures to γ -ray irradiation (which is known to cause DNA double-strand breaks). To reduce the background protein in their experiments, the authors employed the SILAC strategy and only considered proteins enriched in the H2A.X-FLAG sample to be veritable

interaction partners. This resulted in the identification of numerous proteins associated with H2A.X before and after irradiation treatment. An alternative approach to map the interaction partners of histone proteins was utilized by Dirksen et al., which involved attaching peptides with sequences identical to the tail of histone proteins to a solid support and performing immunopurification followed by mass spectrometry (154). This strategy enabled 700 proteins and 90 phosphorylation sites to be identified.

Transcription factors are a general class of proteins that bind to specific DNA sequences and activate (or repress) transcription. Numerous methods have been recently developed to elucidate the specific sequences bound by these proteins; one of the most common techniques used to accomplish this is chromatin immunoprecipitation. This method utilizes a mild formaldehyde treatment to cross-link proteins to DNA followed by DNA shearing and immunoprecipitation of the protein–DNA complexes. After the reversal of the formaldehyde cross-link by heat treatment, the DNA may be analyzed by various means (155). Kim et al. presented an improved method for the detection of the DNA sequences purified with TATA-box binding protein (TBP) in *S. cerevisiae* (156). This method, known as sequence tag analysis of genomic enrichment, or STAGE, was capable of producing specific 21 base pair sequences from the purified DNA and enabled the identification of 79 sequences bound by TBP without the use of microarrays. The authors demonstrated the value of the STAGE method in human cells by mapping the DNA sequence bound by the E2F4 transcription factor. A total of 45 sequences were identified, and 15 out of 18 randomly selected targets were validated by a standard PCR protocol, demonstrating the quality of the methodology. An alternative method termed chromatin immunoprecipitation with paired-end ditag (ChIP-PET) was used to study the DNA sequences bound by p53 (157), c-Myc (158), and two transcription factors, Oct4 and Nanog (159). The ChIP-PET methodology enables the purification and sequencing of 36 base pair DNA fragments containing both ends of the original DNA bound by a bait protein representing an improvement in the amount of information obtained compared to the STAGE strategy.

Microarray technology has become more commonplace in the scientific community, especially in laboratories that investigate gene expression. Microarrays with functionalized surfaces containing oligonucleotides or proteins have been developed in recent years to study the interactions between these two species. Ho et al. created a protein microarray containing 282 known or postulated *S. cerevisiae* transcription factors that were expressed in yeast and purified by GST affinity purification before being spotted on microscope slides (160). The microarray was then probed with 40 Cy3-labeled oligonucleotides containing one or more known sequence motifs. To reduce the experimental noise levels, the arrays were also probed with Cy5-labeled sequences that had two mutated base pairs within the motifs. In total, 211 protein–DNA interactions were observed and a subset of those interactions were confirmed by EMSA. In 2004, Mukherjee et al. probed whole-genome yeast intergenic double-stranded DNA arrays with FLAG tagged Abf1, Rap1, and Mig1, three known *S. cerevisiae* transcription factors (161). Using an anti-FLAG M2 antibody or an anti-GST antibody tagged with Cy3, the DNA bound protein could be identified. The antibody signal was then normalized against the

DNA content of each spot by staining the DNA with Syber Green. This method enabled an extra 75 novel protein–DNA interactions to be identified for Mig1 and 90 and 107 interactions for Rap1 and Abf1, respectively. Berger et al. extended the use of this method by designing a DNA microarray containing all possible 10 base pair sequences in a high-density double-stranded array (162). These arrays were probed with five transcription factors as previously described (161), enabling a number of novel DNA–protein interactions to be identified.

CHEMICAL PROTEOMICS

Interactions between proteins and small molecules are of great interest in drug development and may be used to probe protein function and structure. Godl et al. defined the cellular targets of the small molecule SU6668 by immobilizing the drug on sepharose beads and immunoprecipitating the proteins that interacted with it (163). The interacting proteins were eluted with a concentrated ATP solution, resolved by SDS-PAGE and identified by mass spectrometry. In total, the authors identified 30 proteins interacting with SU6668, 8 of which were kinases not previously known to be affected by SU6668. Kosaka et al. performed a similar study on D942, a drug that increases glucose uptake in L6 myocytes; they immobilized D942 as well as a physiologically inactive derivative to sepharose beads and purified interacting proteins (164). A total of 97 proteins were identified using D942 as the capture agent, 90 of which were also isolated with the inactive derivative. NAD(P)H dehydrogenase I was found to be selectively inhibited by D942, affecting the ATP/AMP ratio in cells. Salisbury et al. developed a chemical affinity reagent based on the structure of a general histone deacetylase inhibitor, suberoylanilide hydroxamic acid (165). The reagent enabled proteins to be cross-linked to interacting proteins following irradiation with UV light and then affinity purified using streptavidin beads. The reagent was used to perform affinity purifications from mammalian melanoma cells (MUM2B and MUM2C), which resulted in the identification of multiple histone deacetylases from both cell lines. They further demonstrated that their affinity reagent could be used to probe histone deacetylase complex composition by showing that while the coREST protein was present in similar levels in MUM2B (aggressive) and MUM2C (nonaggressive) cell lines, coREST was purified to a much higher level from the aggressive cell line, suggesting differences in coREST protein complexes between the cell lines. Fabian et al. performed a systematic screen of 20 kinase inhibitors to define their selectivity toward 440 human kinases (166). They expressed the kinases as fusion proteins using T7 bacteriophages and purified them from *E. coli*. Biotinylated SB202190 was used as a universal bait to capture the various kinases and enable a competitive binding experiment to be performed with the 20 kinase inhibitors. The binding affinity of the kinase inhibitors was determined by quantitative PCR against the T7 bacteriophage or in phage plaque assays. The authors observed that the various kinase inhibitors bound to their intended target with the greatest affinity, however, also interacted with numerous other kinases.

Chemical proteomics has also been used to identify classes of proteins based on their affinity toward specific compounds. For example, Sieber et al. created 3 libraries of small molecules based on an alkyne-tagged hydroxamate–benzophenone probe, resulting

in the synthesis of 60 different compounds with high affinities for metalloproteases (167). The alkyne group was added to each probe to enable the easy addition of rhodamine or biotin tags following its covalent attachment to a target metalloprotease. It was observed that many probes were specific for one or a few metalloproteases. To further increase the enrichment, the authors mixed four probes together and studied various breast cancer and melanoma cell lines. Significant metalloprotease differences were observed between the various cell lines, illustrating the power of their approach. Kumar et al. developed a chemical probe selective for protein tyrosine phosphatase (PTP) labeled with rhodamine (168). The molecule contained an α -bromobenzyl phosphonate, which has been reported to form a covalent attachment with PTPs. The authors confirmed the selectivity of the probe for PTP and not for other proteins through in vitro assays. They utilized their method to define the global activity of PTP in six cancerous mammalian cell lines and observed significant differences between each. They also examined the effect of hydrogen peroxide on PTP and observed a significant decrease in PTP activities following a 15 min treatment.

A popular class of chemical probes in proteomics is protein cross-linkers. They comprise a group of reactive molecules that can covalently link proteins, facilitating the analysis of protein complexes. Suchanek et al. developed a novel cross-linker that consisted of amino acid analogues that could be introduced to proteins in vivo without detrimental cellular effects (169). The cross-linkers were based on the structure of isoleucine, leucine, and methionine and contained a diazirine ring that could be activated by UV light to form a reactive carbene, which covalently linked two proteins in proximity. The authors used the cross-linkers in a proof-of-concept study to study the interaction between PGRMC1, SCAP, and Insig-1 and confirmed that they form a complex together. Recently, Liu et al. expanded this idea and reported an efficient method for the insertion of six unnatural amino acids in mammalian cells with great efficiency and without adverse cellular effects (170).

SYSTEMS BIOLOGY

The study of systems has always been a common element of worldwide research endeavors. Systems investigations are large scale and often require a great deal of monetary resources and time but have the potential to elucidate highly impacting information. Due to these factors as well as others, numerous transdisciplinary laboratories and institutes have recently been established around the world to study biological systems. Systems biology has been defined in numerous ways (171, 172) and has been generally viewed as a broad study of the components of biology, such as proteins and RNA, in an effort to better model their interactions through the use of multiple analytical methods. Here we will review a selection of the major proteomic contributions in the field of systems biology of the last 2 years illustrating various “flavors” of systems biology experiments currently being performed.

To better comprehend the expression and stability of proteins in *S. cerevisiae*, Belle et al. used the yeast TAP-tagged strain collection to determine the protein abundance by Western blot (173). The authors inhibited protein translation with cycloheximide and analyzed the expression of 3751 TAP-tagged proteins

before protein synthesis inhibition as well as after 15 and 45 min of inhibition. The blot intensity data was fitted to a first-order decay function to determine the protein half-life. The mean and median protein half-life observed was determined to be approximately 43 min; 151 of the proteins tested were highly unstable, having a half-life of less than 4 min. To ensure that the TAP tag did not affect the half-life measurements, 24 untagged proteins were tested, revealing no significant differences. By combining their data with previous global reports of messenger RNA abundance and translation rate, the authors gathered evidence that proteins in *S. cerevisiae* are usually observed in one of two systems; one optimized for protein production while the other one favors regulatory efficiency. Sopko et al. systematically analyzed the effect of protein overexpression in the yeast *S. cerevisiae* (174). The authors prepared an array containing 5280 yeast strains with a single gene under the control of an inducible promoter. It was observed that the overexpression of 769 proteins resulted in a reduced growth rate corresponding to approximately 15% of the yeast proteome. Furthermore, 184 of these proteins caused morphological defects while 120 proteins caused cell-cycle arrest. This report demonstrated that yeast cells can easily sustain a 50-fold overexpression of most proteins without experiencing any ill effects, illustrating the robustness of biological systems.

Jones et al. developed a protein microarray containing most SH2 and PTB domains as well as known phosphotyrosine binding domains present in the human genome, which were probed with phosphorylated peptides (175). In total, 159 protein domains were probed with 66 phosphotyrosine peptides corresponding to the known or expected phosphorylation sites of epidermal growth factor receptor (EGFR or ErbB1) and three other EGFR family members. To ensure the observed interactions were due to the presence of phosphorylated tyrosines, the protein microarrays were also probed with nonphosphorylated analogues of the peptides. The protein microarrays were probed with five different peptide concentrations ranging from 10 nM to 5 μ M enabling the determination of the equilibrium dissociation constant (K_D) for each protein domain-peptide pair. The data highlighted that EGFR and ErbB2 interacted with significantly more proteins as their concentration increased while ErbB3 did not. It was postulated that this observation may indicate why EGFR and ErbB2 are highly oncogenic, while ErbB3 is not.

Chu et al. analyzed proteins associated with chromatin in *Caenorhabditis elegans* during spermatogenesis and oogenesis using a conventional chromatin purification method and multidimensional protein identification technology (MudPIT) analysis (176). In total, 1099 proteins were observed with spermatogenic chromatin, 502 of which were found in three independent preparations, and 814 proteins were observed with oogenic chromatin. To identify proteins potentially involved in male fertility, the proteins observed with oogenic chromatin were removed from the data set resulting in 132 candidate proteins to be further tested. RNA interference was used to inhibit the expression of these proteins in *C. elegans*; expression reductions for 50 of the 132 proteins resulted in sterility or embryonic lethality, 19 of which possess known mammalian orthologues. It was further observed that mouse knockouts of 7 of the 19 proteins caused male infertility demonstrating the effectiveness of their experimental design in identifying proteins essential for male fertility.

Stuart et al. reported a detailed analysis of *Drosophila* phagosomes using proteomics data, organelle purification, and analysis as well as previously reported protein–protein interaction data (177). The authors purified phagosomes from *Drosophila* cells using latex beads and determined their protein composition by mass spectrometry, resulting in the identification of 617 proteins. A protein–protein interaction network from five large-scale protein–protein interaction data sets was constructed, and the 617 phagosomal proteins were positioned as anchors within the network. This approach enabled the authors to expand the hypothetical number of phagosomal proteins to 837. Each candidate protein was then knocked down using RNA interference, and the physiological impact on the phagosomes was observed. Approximately 28% of the candidate proteins affected phagocytosis upon knockdown, illustrating the efficacy of the interdisciplinary experiment.

PROTEIN BIOMARKERS

The proteomic community has paid increasing attention to the need for biomarker discovery. The applications of biomarkers are broad and range from diagnosis and prognosis of diseases to the monitoring of a biological response or therapeutic intervention. The poor success rate in introducing new drugs into the market combined with recent re-evaluations of approved drugs has further reinforced the need for quality biomarkers. The primary reason for this poor success rate is due to unexpected toxicity or side effects in late-phase clinical trials. Furthermore, some well-publicized cases of approved drugs being taken off the market due to side effects have been observed in recent years. This obviously represents a serious health risk and comes at tremendous financial cost. To address these critical issues, the FDA has proposed several new recommendations (178) among which is the use of biomarkers to aid in drug candidate selection, attrition, optimization, and confirmation. It is therefore quite obvious that proteomics technologies are going to play a critical role in the discovery of biomarkers now and for years to come.

The application of proteomic strategies in biomarker discovery grew tremendously in 2005–2006; 67% of all papers reported in PubMed with the words “proteomic” and “biomarker” in their title or abstract were released in 2005–2006. From 2002 onward, the number of papers published per year in this specific field has increased linearly (r^2 of 0.99). A third of the papers in 2005–2006 are reviews, a fifth report results of surface-enhanced laser desorption/ionization (SELDI)-based approaches, while the remainders either contain different applications of proteomics for biomarker discovery or report on technical developments to improve biomarker discovery. Here we have decided to focus on the challenges that the proteomic biomarker discovery field is facing as well as some of the potential solutions reported in the literature during 2005–2006.

Different approaches have been designed and implemented for the discovery of protein biomarkers. The bulk of the approaches thus far study serum/plasma as sources of potential biomarkers, while the remainder primarily focus on discovering potential biomarkers in tissues and cell lines followed by validation of the presence of these markers in blood. The focus of this section will be on biomarker discovery from plasma and serum.

Collection and Handling of Serum and Plasma. The collection and handling of serum and plasma prior to their analysis by proteomic approaches can adversely affect the outcome of the investigation. Unfortunately, no clear guidelines are available for proper collection and handling of serum and plasma, which is further aggravated by the lack of a gold standard available for serum and plasma experiments. This makes the comparison of different analytical techniques encountered in the literature for serum and plasma analysis difficult and likely biased due to different sample handling approaches.

Recently, West-Nielsen et al. (179) used principal components analysis (PCA), support vector machines, and clustering based on genetic algorithms to compare the impact of different serum handling techniques on the outcome of proteomic analysis. They concluded that clotting time and temperature are key contributors for serum quality. In particular, blood left to clot at room temperature for more than 3 h or more than 24 h at 4 °C impacted the proteomic results. Similarly, storage of sera beyond 4 h at room temperature or 24 h at 4 °C negatively impacted the proteomic results.

Hsieh et al. (180) found that blood collecting tubes, anticoagulants, length of time before centrifugation for serum/plasma isolation, and length of time before hemolysis impacted the plasma and serum proteome in the low-molecular-weight region (below 25 000). They also observed minimal impact for the storage of samples at 4 °C up to 24 h in agreement with West-Nielsen (179). Rai et al. (181) performed a detailed study of the factors that influence the outcome of plasma and serum proteomic analysis. They concluded that plasma might be easier to control over serum for proteomic analysis; however, they indicated that the sample handling methodology needs to be adjusted according to the analytical technique.

Villanueva et al. (182) showed that sample handling can lead to false positive biomarker discovery. For example, they found “biomarker patterns” with diagnostic accuracies of 100% when comparing the mass spectrometry data from two identical sets of control blood samples prepared in different tubes. Similarly to Hsieh (180) and West-Nielsen (179), they found that the time allowed for clotting affected the outcome of the analysis. However, contrary to Hsieh (180) and West-Nielsen (179), they found that freeze/thaw cycles affected the results.

Protein Depletion from Plasma and Serum. The discovery of protein biomarkers in plasma/serum faces some serious challenges. The dynamic range of protein concentrations is estimated to cover 8 orders of magnitude; to date, no analytical technique can cover such a large dynamic range. However, 98% of the protein mass present in serum is composed of approximately 22 proteins (183). Hence, strategies to remove these proteins from plasma/serum have been proposed, and recently, depletion kits that are primarily immobilized antibodies against high-abundance proteins have been introduced.

Solubility-Based Approaches. Colantonio et al. (184) reported that albumin can be efficiently removed from serum using a method termed depletion of albumin component (DAC). The DAC method is a multistep protocol consisting of selective precipitation of non-albumin proteins using salt, ethanol, and centrifugation. Although their gels showed a marked decrease in albumin in the pelleted fractions, the amount of depletion remains

unclear. Based on their results, it seems that merely depleting albumin is insufficient for efficient biomarker discovery since immunoglobulin concentrations remain elevated. Chen et al. (185) presented a modified TCA/acetone protocol, which they claimed provides 60% removal of albumin. They compared their TCA/acetone approach with the SwellGel Blue Albumin Removal Kit and the MontageAlbumin Deplete Kit and obtained respectively 60, 80, and 50%, of albumin depletion. Although their TCA/acetone approach did not deplete the greatest quantity of albumin, they claimed it was the most specific to albumin. In general, solubility-based depletion approaches are simple and inexpensive methods of decreasing albumin concentrations in plasma and serum; however, they are not as efficient as commercially available antibody methods.

Antibody-Based Approaches. Agilent commercialized the Multiple Affinity Removal System (MARS) strategy, an immunoglobulin G (IgG)-antibody-based immunoaffinity LC column that reduces the level of six of the most abundant proteins (186) in plasma/serum (albumin, IgG, immunoglobulin A (IgA), haptoglobin, transferrin, α -1-antitrypsin). Zolotarjova et al. (187) compared the performance of the Cibacron Blue (CB) dye method and the Agilent MARS column and cartridges for the depletion of high-abundance proteins from serum. The CB method has been shown to bind albumin (188) but has also been previously shown to bind other proteins as well. (189) Zolotarjova et al. (187) analyzed the proteins that flow through the CB and MARS columns as well as those that are bound. They found that the CB depletion of albumin was poor and that 60 proteins were bound by the CB resulting in a low specificity. In contrast, the MARS column approach showed 98% efficiency in removing albumin, transferrin, α -1 antitrypsin, and haptoglobin. A similar study of five depletion methods was reported by Bjorhall et al. (190) concluding that the Agilent MARS column approach was the most efficient with a 99.4% removal of albumin and depletion of other high-abundance plasma proteins. A close second was the Albumin and IgG Removal Kit from Amersham Biosciences, which removed 99.5% of albumin. Echan et al. (191) also studied the performances of the Agilent MARS depletion system and obtained similar results. Yocum et al. (192) also obtained good depletion of protein using the MARS column. However, they found that 24 proteins that were present in their serum were not observed in the depleted sample, indicating a potential nonspecific depletion of the plasma proteome. It appears that the antibody-based approaches are very efficient at removing the targeted proteins, i.e., high-abundance plasma proteins. However, the literature remains ambiguous regarding the impact of the antibody-based approaches on other proteins.

Huang et al. (193) developed the GenWay Seppro IgY12 column (also sold by Beckman-Coulter under the name ProteomeLab IgY12) for the depletion of 12 high-abundance proteins from human plasma (albumin, IgG, α -1-antitrypsin, IgA, IgM, transferrin, haptoglobin, α -1-acid glycoprotein, α -2-macroglobulin, apolipoprotein A-I, apolipoprotein A-II, fibrinogen). These 12 proteins represent 95% of the total protein mass in human plasma. The column seemed to perform well for the depletion of these proteins from human, even though no percentages of depletion were reported. Liu et al. (194) studied the performance of this system for the depletion of 12 high-abundance proteins from

human plasma in greater detail and demonstrated that the performance of the IgY-12 column is reproducible. However, over 30 proteins other than the 12 listed above were also shown to bind to the IgY-12 column. The ProteomeLab system also seems to perform very well for the removal of the top 12 proteins from plasma. However, as with other available systems, off-target protein depletion seems to be occurring.

The depletion of the top 6 proteins on the Agilent MARS column or the top 12 plasma proteins on the Beckman ProteomeLab IgY12 column from human plasma/serum using antibody-based approaches seems to work reasonably well and can reduce the total protein mass respectively by 85 and 95%. Gong et al. (195) showed that the combination of these two approaches and the analysis of the flow-through and bound fractions lead to a higher number of identified proteins. Interestingly, they observed an approximately 60% overlap in the proteins identified in the bound fractions and in the flow-through fractions between the Agilent and the Beckman systems. They observed a large number of proteins bound to the columns, which they attributed to the carrier effects of albumin and other high-abundance plasma proteins. Although these two approaches are a first step to enhance our ability to study the complexity of plasma/serum, it is likely that further depleting of these fluids will be critical. However, the cost of removing additional proteins by antibodies might become prohibitive. Furthermore, these depletion approaches need to be coupled with efficient protein/peptide separation techniques to maximize their effects.

Animal models for human diseases are used extensively to better comprehend the processes involved in diseases and to study the impact of drugs. Animal models can be used for the discovery of biomarkers; however, the depletion techniques used in human samples might not be as efficient when animal samples are used, depending on the orthology of the proteins. Duan et al. (196) demonstrated that albumin can be depleted from mouse plasma using goat anti-mouse albumin antibody immobilized on agarose beads. 2D gel electrophoresis of mouse plasma depleted using this approach failed to show the presence of albumin in the sample. Freeman et al. (197) tested the Agilent MARS column for the depletion of monkey serum. Although it was not completely depleted of albumin, transferrin, IgG, IgA, α -1-antitrypsin, and haptoglobin using this approach, the depleted serum had 83% less protein than the whole serum therefore making it a reasonable approach to improve the study of monkey serum. Mice, rats, and to a lesser extent monkeys and dogs are used as animal models, and depletion approaches developed for human plasma/serum will perform differently for these models and need to be further investigated.

The plethora of sample depletion techniques make it difficult to rapidly select a method to incorporate in a biomarker discovery process. Fortunately, Whiteaker et al. (198) recently performed a comparative study of eight popular serum fractionation approaches while the rest of the platform remained constant. They compared methods based on: enrichment of cysteinyl-peptide (199) and N-linked glycopeptides (94), fractionation methods based on magnetic bead separation (200) and size fractionation, and finally depletion methods based on protein A/G depletion and immunoaffinity (186). They concluded that depletion methods based on immunoaffinity approaches provided the highest pro-

teome coverage and reproducibility. Furthermore, they postulated that a combination of immunoaffinity depletion followed by cysteinyl-peptide enrichment might provide an even better coverage. Their experiments also demonstrated that, in many cases, the most popular serum fractionation approaches in combination with LC-MS/MS might not provide sufficient coverage of the plasma proteome for biomarker discovery.

ANALYTICAL TECHNIQUES

Lab-on-a-Chip. Lab-on-a-chip components have been developed and used specifically for the rapid analysis of biomarkers. Although most are still in the technical development stage, it is thought that microfluidic devices will provide multiplexing of samples and integration of multiple steps resulting in improved proteomic analysis.

Agilent recently introduced a microfabricated HPLC-chip system that can be interfaced to ESI-MS (201). Hardouin et al. (202) studied the analytical performances of the Agilent HPLC-Chip-MS system. They reported that the microfabricated HPLC-chip can handle five times less concentrated sample than the conventional nanoLC-MS. The potential utility of the HPLC-Chip-MS system for the analysis of plasma was demonstrated by Fortier et al. (203). The analytical performance of the system was measured for a complex protein mixture of 626 different peptides (eight protein digest) over 10 replicate injections. The RSD values for the retention time, m/z , and intensity were respectively less than 0.5, 0.003, and 0.2%. Very good results were obtained in spike experiments in plasma for which as low as 2.4 fmol of spiked protein digests were detected in 100 ng of plasma using the device. Moreover, the addition of a strong cation-exchange trapping column to the HPLC-Chip-MS system allowed proteins from larger volume of plasma to be concentrated and analyzed on the device. For example, insulin-like growth factor binding protein 1 present at less than 80 ng/mL in plasma was detected using this combined approach.

Lazar et al. (204) described an LC system on a chip driven using a multichannel electroosmotic flow pumping technique. The device was coupled to an ESI-MS/MS mass spectrometer for the analysis of complex proteomes obtained from cell lines for biomarker discovery. Very good analytical performances (45 000–180 000 plates/channel) were obtained on this device, although the loading capacity was lower than a conventional nanoLC-MS system. They demonstrated the potential of this device for the analysis of complex proteomes using MCF7 cell lysate separated by strong cation-exchange chromatography.

Herr et al. (205) reported the integration of multiple sample treatments for the analysis of MMP-8 in saliva, a potential marker of periodontic disease. The device consisted of three regions filled with different polyacrylamide gels. At high pore size, the polyacrylamide gel allowed the samples (saliva proteins and fluorescently labeled antibodies against MMP-8) to be driven by electrophoretic mobility. The proteins moved through the high pore size polyacrylamide and encountered a very low pore-sized polyacrylamide gel at a T-cross. At that point, only the small ions could go through and the proteins and antibody accumulated at the junction. After an incubation period, the analytes were mobilized in a separating polyacrylamide gel through another branch of the T junction, and the complexes of fluorescently

labeled monoclonal antibody and MMP-8 were detected by laser-induced fluorescence. The potential of this device was demonstrated on a group of patients preclassified as clinically healthy or periodontally diseased.

Yang et al. (206) reported an interesting application of microfabrication for the continuous separation of plasma from blood. The device consists of an inlet for the blood and a main channel that leads to the blood cell outlet. Perpendicular to the main channel is a series of smaller channels that lead to the plasma outlet. In this device, the blood cells have to travel into the higher flow rate main channel while no cells travel in the lower flow rate perpendicular channels as long as the flow rate ratio is higher than 2.5 and the cell-to-vessel diameter ratio is approximately 1 (Zweifach-Fung effect). The plasma selectivity obtained from their experiments was close to 100% (i.e., no blood cells found in the plasma outlet). This may well represent an excellent technique to prepare plasma samples using a microfluidic device. One caveat, however, is that the plasma outlet was not tested for the presence of protein from lysed cells.

Overall, the application of lab-on-a-chip devices to the field of biomarker analysis from plasma/serum was timid. Presently, this field is still developing different strategies to handle complex samples, and the integration of these components on-chip to address the complexity of blood remains to be developed.

Free-Flow Electrophoresis. Free-flow-electrophoresis (FFE) was introduced as an analytical technique 35 years ago (207). The coupling of FFE and mass spectrometry, however, is relatively recent (208). In the last 2 years, the use of FFE for the analysis of plasma proteins has been reported.

Moritz et al. (209) described a partial proteomic analysis of a plasma sample from the Human Proteome Organization (HUPO) Plasma Proteome Project using FFE followed by mass spectrometric identification of the proteins in some of the FFE fractions. Briefly, their approach integrated the Agilent MARS depletion followed by 2D liquid-based FFE and an HPLC approach (210). FFE was used to separate the plasma proteins by isoelectric focusing into 96 fractions, each corresponding to 0.02–0.10 pH unit. Approximately 16% of the FFE fractions were then separated by HPLC followed by mass spectrometry. Their partial analysis of the FFE fractions (16%) led to the unambiguous identification of 78 proteins, of which 55 were based on at least 2 tryptic peptides. Cho et al. also used FFE to study human serum from the HUPO Plasma Proteome Project. Their strategy focused on the depletion of high-abundance proteins using the Agilent MARS system, fractionation of the depleted plasma proteins by FFE followed by 1D gel electrophoresis of the fractions. One out of the 44 fractions separated by 1D gel electrophoresis was further analyzed by HPLC-ESI-MS/MS leading to the identification of 39 proteins. These preliminary studies illustrate the potential application of FFE for the study of plasma and serum. However, a full study of the plasma proteome analyzed by FFE coupled to mass spectrometry remains to be published.

Application of Protein Arrays To Study Plasma and Serum. The last 2 years saw extensive development in the field of protein arrays. These are described above; however, this section will cover the application of protein arrays for the study of plasma and serum.

Sanchez-Carbayo et al. (211) designed an antibody array for the analysis of bladder cancer samples that used antibodies to recognize genes that were previously demonstrated to be differentially expressed in bladder tumors. This protein array was then used to screen serum samples from different patients. Their array contained 254 antibodies and correctly discriminated bladder cancer patients from controls ($n = 95$) with a correct classification rate of 93.7%. Shafer et al. (212) also developed an antibody array for prostate cancer biomarker discovery. Their array was used to screen serum from patients to discover potential biomarkers. They found a set of proteins that changed in diseased versus normal patients, the most significant being thrombospondin-1, which was strongly elevated in patients with benign prostatic disease and repressed in patients with prostate cancer.

Horn et al. (213) used a protein array consisting of 37 200 redundant recombinant human proteins to probe for elements responsive to the IgG autoantibody repertoire from patients suffering from dilated cardiomyopathy as well as from control patients. Interestingly, a set of proteins known to be associated with heart failure was responsive. Further validation experiments are obviously needed, but the approach seems promising. Andersen et al. (214) developed a conceptually similar approach that was based instead on 626 different peptides in an array that was then used to probe the immunoglobulin complement of serum. Gannot et al. (215) developed a technique called a layered peptide array, in which each layer on a stack of blot transfers represented one peptide for specific antibodies. As antibodies passed through the layers, they preferentially bound to a specific layer that had the corresponding epitope. Multiple serum samples can be screened simultaneously through the device. These papers report technical developments and proof-of-principle experiments that are very exciting and will need to be followed up with a series of systematic studies.

Data Handling. The handling and statistical analysis of data generated from large-scale proteomic studies of plasma and serum is a serious issue in the literature. The most controversial aspect has been the SELDI and MALDI types of analyses in plasma and serum for biomarker discovery. Although the discussion on data handling and statistical analysis can be extended to the bulk of the proteomic techniques used to analyze plasma and serum, we have opted to focus on biomarker discovery by SELDI and MALDI with the expectation that the lessons learned from these will be applicable to other techniques.

Villanueva et al. (182) showed that, in MALDI-MS identification of biomarkers, the algorithm used to find peaks in the spectra and discrepancies in the calibration of the instrument between experiments can lead to different peak selection by the PCA and likely identify false positive biomarkers. It is noteworthy that previous data published using this approach (216) has generated intense discussion and criticism (217).

SELDI-MS has been used extensively for the discovery of biomarkers due to its ease of use and relatively low cost. Unfortunately however, the technique remains very contentious in the scientific community. In particular, recent reports have highlighted the lack of reproducibility in biomarker discovery projects (218) and the lack of sensitivity of the technique (219, 220). Rai et al. (221) analyzed some of the results obtained by SELDI from the HUPO plasma proteome project. Their findings

clearly highlighted inconsistencies from lab to lab, in particular in sample handling. They were able to find peaks that distinctly separated the different samples; however, these peaks appeared to be associated with the most abundant proteins in the HUPO samples. Therefore, SELDI-MS seems to successfully find distinct features between samples; however, the results are greatly influenced by sample preparation and by the lack of protein fractionation. A recent paper by Diamandis et al. (217) on the evaluation of peptidomic and proteomic biomarkers raised key questions regarding the overall validity of these approaches for biomarker discovery and questioned their ability to stand up in clinical application.

Overall we can conclude that the application of proteomic technologies in biomarker discovery needs to do the following: (1) be carefully designed to avoid pitfalls associated with overfitting the data, (2) preferably be validated through double blind studies, (3) be reproducible in different laboratories, and (4) preferably have the results validated by alternative methodologies.

BIOINFORMATICS FOR PROTEOMICS

In the past 2 years, the development of bioinformatic tools for proteomic applications has focused on ensuring that the life sciences community has access to more open source software as well as better statistical methods and filtering algorithms to validate proteomics data.

Open Source Projects. The open source license has changed not only the world of informatics but also the field of proteomics and life sciences. Today, there are numerous umbrellas devoted to hosting open source projects; arguably the most known are Bioinformatics.org (over 300 projects), sourceforge.net (over 900 projects), and Open Bioinformatics Foundation (hosts toolkits like BioJava, BioPerl, and BioPython) (222). Bioinformaticians also have access to the optimized Linux operating system, BioLinux, which is configured specifically for the bioinformatics needs of speed, stability, and security. Under the BioLinux umbrella, we can find complete system projects like BioBrew (bioinformatic-s.org/biobrew/), BioLand (hwww.bioland.cbi.pku.edu.cn/), BioLinuxBR (www.biolinux.df.ibilce.unesp.br/), and Debian-Med (www.debian.org/devel/debian-med/) and software repositories like BioLinux and BIORpms for Red Hat/Fedora (223). One of the last distributions available is NEBC Bio-Linux developed by the NERC Environmental Bioinformatics Centre (223). Unlike the other BioLinux distributions, the NEBC Bio-Linux is an image of a computer that is on the NEBC facility. The advantage of this kind of installation is that the NEBC team can easily troubleshoot the user system and provide better downstream support. Spjuth et al. developed an open source software application call Bioclipse that provides a complete and extensible tool kit for cheminformatics and bioinformatics (224). This java-based Rich Client Platform (RCP) software integrates multiple frameworks and components like Biojava, Chemistry Development Kit (CDK), JChemPaint, Spectrum, and many more. Bioclipse is an advanced powerful workbench that offers resources for (i) molecules, proteins, sequences, and spectra, (ii) 2D editing and 3D visualization, (iii) networked servers, clusters and databases, and (iv) local file systems and devices (e.g., printers). This workbench beautifully integrates cheminformatics and bioinformatics and is based

on RCP, making it easier to extend it to other plug-ins developed in-house.

Quantitative Software Development. The development of quantitative proteomics technology has become an important issue for large-scale proteomic data and was briefly mentioned above. Saito et al. introduced AYUMS, an automatic console-based software program customized specifically for LC–MS/MS proteomic data obtained by the SILAC technique (225). Compared to other quantitative software, it is not restricted to the analysis of LC–MS data only (while MZmine (226) is), does not require a large amount of memory, and is platform-independent (which is not the case for MSQuant (227)). The AYUMS software automatically calculates peak ratios for differentially labeled peptides directly from the raw MS/MS data file and generates an output report in CSV format. In 2004, the iTRAQ quantitative technique was introduced (228), and until recently, only the ProQuant software, developed by Applied Biosciences, was capable of performing quantitative analysis of MS/MS data with iTRAQ labeling. To remedy to this situation, i-Tracker and a software package called Multi-Q were recently introduced for the analysis of iTRAQ labeled MS/MS data. i-Tracker, created by Shadforth et al. (229), calculates iTRAQ reporter ion ratios and reports them in CSV format, which can then be integrated with Mascot or Sequest results. Shadforth and his colleagues were able to demonstrate that i-Tracker computes nearly identical output results compared to the ProQuant software in terms of calculating the ratios of reporter ion peaks. i-Tracker is written in Perl, is platform-independent, and can be freely downloaded. Hence, i-Tracker is an excellent alternative for the analysis of iTRAQ labeled samples. The software package, Multi-Q, was developed by Lin et al. (230) and provides converters for spectral data files from major MS manufacturers, WIFF (Applied Biosystems), RAW (Thermo Finnigan and Walters) and BAF (Bruker Daltonics), and reduces them to a mzXML file. In addition, Multi-Q accepts search results from different engines (MASCOT, SEQUEST, X!Tandem) in CVS or HTML formats. Moreover, Multi-Q allows the user to define their own filtering and statistical module to correct: (i) low-quality MS/MS spectra; (ii) random and systematic errors in detection responses; (iii) sample handling.

Data Analysis, Data Management, and Data Exchange Standards. Proteomics experiments tend to generate a significant amount of data, and publications rarely include the raw data. Furthermore, the “massaged” final data sets are often buried in the Supporting Information, which are not readily accessible through search engines or disappear rather quickly from some journals. There have been many discussions in the community on the need for more transparency in proteomic publications. Different software, databases, and web sites have been introduced for the curation of proteomic data. However, in our view, these efforts are still work in progress. The submission of proteomic data through the different approaches reported to date is often tedious and requires a level of expertise that might become a barrier for new investigators in the field. There is a tremendous need for simple interfaces for data submission, stable funding for proteomic curation sites ensuring that data submitted today will be available in the foreseeable future, and clear annotation of the different levels of information (raw vs final) to ensure that public proteomic results do not confuse the broader biology community.

In 2005, Garden et al. introduced a new open source client-server, called PROTEIOS, for the management of biomaterial information, raw data, images, and analysis results (231). In addition, PROTEIOS provides additional plug-ins for protein identification, data viewing, and tool analysis. Furthermore, this application can import mzData, mzXML, and sample generation/sample processing parts of PEDRo (Proteome Experimental Data Repository) (232). PROTEIOS is implemented in java, SQL, and is platform-independent. Rauch et al. have recently introduced an open source Computational Proteomics Analysis System (CPAS) (233). In contrast to other software management systems (such as SBEAMS and PRIME), CPAS includes the following: (i) multiple standard file formats such as FASTA, mzXML, and pepXML; (ii) an Experiment Annotation to track and organize biological experiments and view the workflow; (iii) tools to promote collaborative projects. Cerami et al. have developed cPath, an open source database and web application to collect, store, and query biological pathway data (234). This web-based software can import pathway formats from multiple databases (DIP, MINT, KEGG, Reactome, HPRD, IntAct, and many more), visualize the pathway, and export it to a third party software program like Cytoscape (a software visualizing platform for molecular interaction networks) (235). Furthermore, cPath implemented plug-ins to support PSI-MI (236) and BioPAX (www.biopax.org). PSI-MI and BioPAX are two different standards for the representation and exchange of pathway data. PSI-MI is an initiative of the HUPO and is based on the XML format. The data found in the PSI-MI standard are structured around an entry tag. The entry tag can represent one or more interactions that are grouped together. The pathway is described between the interaction List tags. The BioPAX standard is an initiative of a group of collaborators who are trying to define a unified framework for sharing pathways. BioPAX is defined in different levels, where level 1 represents the metabolic networks and level 2 represents the molecular interaction networks.

The XML is a language that facilitates the sharing of data across different information systems such as the Internet. Stanislaus and his colleagues have developed an Annotated Gel Markup Language (AGML) Central, a web-based open source software program for the handling of two-dimensional gel electrophoresis data in the AGML format (237). The AGML format provides standardization for the 2D gel electrophoresis (2-DE) and for storing MS data related to specific spots. Moreover, AGML defines the minimum information required for the experimental settings and experimental results. The AGML Central implemented converters for other formats (like PDQUEST, Phoretix, and Melanie) and a search tool to visualize other 2-DE databases such as SWISS-2DPAGE and SIENA-2DPAGE. Recently, Le Novère et al. have introduced standards for curating quantitative models of biological systems called Minimum Information Requested in the Annotation of Biochemical Models (MIRIAM) (238). MIRIAM established a set of rules that described how biochemical models should be structured by using the markup language and a detailed annotation scheme. This annotation scheme uses a unique identifier that links to the full description of the models.

Interaction Networks, MS Identification, and Databases. The mapping of protein–protein interaction has been a highly productive segment of proteomics in 2005–2006 (239), but

recently (240, 241), the quality of these data sets has been questioned due to high false positive levels. APID is a web-based software program that integrates and analyzes protein interactions from five main source databases (BIND, DIP, HPRD, IntAct, MINT) (239). One of the unique attributes of this software are several score parameters that weigh the reliability and the functional meaning of the protein–protein interaction. The possible score parameters are divided in two sections: (i) the proteins where the parameters are the connectivity, the cluster coefficient, the gene ontology (GO) functional, and enrichment; (ii) the interactions where the parameters are the numbers of methods, the GO overlapping and the iPfam domain–domain interaction. To unify the data from all of the databases, APID use three key reference identifiers: (i) Uniprot ID to link the protein to its sequence and protein information; (ii) PSI-MI ID to unify the experimental methods; (iii) PubMed ID to validate the experimental methods to a publication. In addition, APID includes a tool to visualize and navigate through subnetwork interaction. Rinner et al. describe a new approach to analyze protein networks using mass spectrometry and computational analysis (242). This approach consists of the alignment of LC–MS patterns from the bait and the control coimmunoprecipitation (Co-IP) and clustering the MS1 signal. The cluster contains peptides from the protein that specifically purifies with the bait. The peptides are then assembled into proteins based on high-confidence MS2 spectra (obtained by PeptideProphet) using the SuperHirn software. Compared to the conventional Co-IP technique, this approach improves the sensitivity and the specificity in the identification of protein–protein interactions. One danger of solely mining databases to determine protein interactions is that the outcome of the study is dependent on the quality of the curation. It appears that some of the protein interaction databases do not differentiate interactions obtained by yeast-two-hybrid, immunoprecipitation coupled to mass spectrometry, and even genetic interactions. Although those are all called interactions, they do not represent identical biological situations.

Over the years, a rising interest has grown within the bioinformatic community to statistically evaluate the ability of existing software packages to correctly match theoretical and observed mass spectra. Mascot (243), OMSSA (244), RMET-Luck (245), and PepProbe (246) are some of the software programs that are commonly used for peptide and protein identification. Tabb et al. has introduced MyriMatch, a database search algorithm where the calculation of peptide score is based on a multivariate hypergeometric distribution (247). The use of this type of scoring places a greater emphasis on matching the intense peaks rather than all the peaks. Compared to Sequest and X!Tandem, the MyriMatch search engine demonstrates a better discrimination of scoring. These findings indicate that the lack of differentiation in the intensity of the peaks limits the existing search engines' ability to score a peptide match. Frequently, in LC–MS/MS experiments, proteins are identified by a single peptide match. Higdon et al. has introduced a method to differentiate between true and false identifications among single-peptide match (248) by doing the following: (i) estimating the rate of false matches using a randomized database search; (ii) using a logistic regression model to identify and assign probabilities for single hit; (iii) removing possible biased matches. This

method results in a 22–65% increase in the identified proteins with an error rate less than 2%.

Titulaer and colleagues have introduced a new database application for detecting and identifying peptide expression in protein profiles obtained from control groups and bodily fluids from patients (249). This java GUI software integrates (i) a connection to the software package R for the statistical calculation, (ii) a connection to the MySQL database that contains all the metadata, and (iii) a connection to an FTP server to store raw MS fid (free induction decay) files and processed data. The metadata of a sample contains all the information for the sample such as its origin, the experiment performed on the sample, and a link to clinical information. The database application uses the Wilcoxon–Mann–Whitney statistical test from R to compare the peptide profile from the patient to the control group. Titulaer et al. demonstrated that the Wilcoxon–Mann–Whitney statistical test can be used to distinguish a peptide profile from breast cancer patients with leptomeningeal metastases and a prostate cancer patient. The use of multiple protein identifiers for the same protein has become a general problem in the proteomic field. Major public databases try to eliminate this problem by doing an exhaustive cross-reference, but the maintenance of that kind of structure is time-consuming. Babnigg et al. have introduced a database of unique protein sequence identifiers called Sequence Globally Unique Identifiers (SEGUID) based on the secure hash algorithm (SHA-1) digest of the primary protein sequences (250). Therefore, the use of these identifiers can serve as links to different databases and ensure that proteomics data is resilient to change in the annotation databases.

CONCLUDING REMARKS

The field of proteomics has had its share of success stories in 2005–2006. A few examples would be protein interaction mapping and the application of quantitative proteomics. Nevertheless, the field has also had challenges such as the quality of data interpretation in biomarker discovery, which has created a wave effect across the whole field of proteomics.

We predict that the development of parallel technological approaches, such as protein arrays, will become increasingly important to the future development and progress of the proteomics field. Additionally, functional proteomics and methods for probing subproteomes will be strong foci in the coming years. We also predict that open access to proteomic data will become the norm, which will provide better transparency and will ensure that the submission of proteomic data sets to public databases is made simple. Beyond creating repositories to house the massive amounts of raw and finalized proteomic data sets, it will be important to have appropriate and simple mining tools that follow universally accepted standards. Otherwise, our efforts will risk confusing the bulk of the biology community and increase the skepticism surrounding the field of proteomics.

ACKNOWLEDGMENT

J.C.S. and J.-P.L. are co-first authors of this review.

Jean-Philippe Lambert completed his Honors B.Sc. degree in biochemistry at Concordia University, Montréal. He is currently a Ph.D. candidate at the University of Ottawa in biochemistry and is a member of the Ottawa Institute of Systems Biology. His current research interests are in the characterization of protein complexes associated with chromatin using novel proteomics tools and mass spectrometry.

Jeffrey C. Smith completed his Honors B.Sc. degree in biochemistry at Trent University in 2000. He received his Ph.D. degree in chemistry from York University at the Centre for Research in Mass Spectrometry in the spring of 2005. During his Ph.D. studies he was recognized as an Ontario graduate scholar, as well as an NSERC scholar. He is presently working as a postdoctoral fellow for Professor Daniel Figeys at the Ottawa Institute of Systems Biology and is developing methods to study the dynamics of protein phosphorylation in a variety of biological systems.

Fred Elisma completed a B.Sc. degree in biochemistry, M.Sc. degree in biology, and a Diplôme d'Études Supérieures Spécialisées degree in bioinformatics at the Université du Québec à Montréal. He is currently a bioinformatician for Professor Daniel Figeys at the Ottawa Institute of Systems Biology.

Daniel Figeys is a professor in the Department of Biochemistry, the Director of the Ottawa Institute of Systems Biology, and a Tier-1 Canada Research Chair in proteomics and systems biology. Daniel obtained a B.Sc. and a M.Sc. in chemistry from the Université de Montréal. He obtained a Ph.D. in chemistry from the University of Alberta and did his postdoctoral studies at the University of Washington. Prior to his current position, Daniel was Senior VP of Systems Biology with MDS-Proteomics. From 1998 to 2000, he was a Research Officer at the NRC-Canada. Daniel's research involves developing proteomics technology and their applications in systems biology.

LITERATURE CITED

- Lambert, J. P.; Ethier, M.; Smith, J. C.; Figeys, D. *Anal. Chem.* **2005**, *77* (12), 3771–87.
- Chapman, T. *Nature* **2005**, *434* (7034), 795–8.
- Lichty, J. J.; Malecki, J. L.; Agnew, H. D.; Michelson-Horowitz, D. J.; Tan, S. *Protein Expression Purif.* **2005**, *41* (1), 98–105.
- Ethier, M.; Lambert, J.-P.; Vasilescu, J.; Figeys, D. *Anal. Chim. Acta* **2006**, *564*, 10–18.
- Gingras, A. C.; Aebersold, R.; Raught, B. *J. Physiol.* **2005**, *563* (Pt. 1), 11–21.
- Monti, M.; Orru, S.; Pagnozzi, D.; Pucci, P. *Biosci. Rep.* **2005**, *25* (1–2), 45–56.
- DeVoe, D. L.; Lee, C. S. *Electrophoresis* **2006**, *27* (18), 3559–68.
- Lazar, I. M.; Grym, J.; Foret, F. *Mass Spectrom. Rev.* **2006**, *25* (4), 573–94.
- Sung, W. C.; Makamba, H.; Chen, S. H. *Electrophoresis* **2005**, *26* (9), 1783–91.
- Miao, J.; Wu, W.; Spielmann, T.; Belfort, M.; Derbyshire, V.; Belfort, G. *Lab Chip* **2005**, *5* (3), 248–53.
- Wheeler, A. R.; Moon, H.; Bird, C. A.; Loo, R. R.; Kim, C. J.; Loo, J. A.; Garrell, R. L. *Anal. Chem.* **2005**, *77* (2), 534–40.
- Ethier, M.; Hou, W.; Duetwel, H. S.; Figeys, D. *J. Proteome Res.* **2006**, *5* (10), 2754–9.
- Vasilescu, J.; Zweitzig, D. R.; Denis, N. J.; Smith, J. C.; Ethier, M.; Haines, D. S.; Figeys, D. *J. Proteome Res.* **2007**, *6* (1), 298–305.
- Hou, W.; Ethier, M.; Smith, J. C.; Sheng, Y.; Figeys, D. *Anal. Chem.* **2007**, *79* (1), 39–44.
- Balboni, I.; Chan, S. M.; Kattah, M.; Tenenbaum, J. D.; Butte, A. J.; Utz, P. J. *Annu. Rev. Immunol.* **2006**, *24*, 391–418.
- Calvo, A.; Gonzalez-Moreno, O.; Yoon, C. Y.; Huh, J. I.; Desai, K.; Nguyen, Q. T.; Green, J. E. *Mutat. Res.* **2005**, *576* (1–2), 66–79.
- Hudelist, G.; Singer, C. F.; Kubista, E.; Czerwenka, K. *Anticancer Drugs* **2005**, *16* (7), 683–9.
- LaBaer, J.; Ramachandran, N. *Curr. Opin. Chem. Biol.* **2005**, *9* (1), 14–9.
- Le Guezennec, X.; Brinkman, A. B.; Vermeulen, M.; Denissov, S. G.; Gazzola, C.; Lohrum, M. E.; Stunnenberg, H. G. *BJU Int.* **2005**, *96* Suppl 2, 16–22.
- Lueking, A.; Cahill, D. J.; Mullner, S. *Drug Discovery Today* **2005**, *10* (11), 789–94.
- Maercker, C. *Biosci. Rep.* **2005**, *25* (1–2), 57–70.
- Ramachandran, N.; Larson, D. N.; Stark, P. R.; Hainsworth, E.; LaBaer, J. *FEBS J.* **2005**, *272* (21), 5412–25.
- Sakanyan, V. J. *Chromatogr., B: Anal. Technol. Biomed. Life Sci.* **2005**, *815* (1–2), 77–95.
- Sheehan, K. M.; Calvert, V. S.; Kay, E. W.; Lu, Y.; Fishman, D.; Espina, V.; Aquino, J.; Speer, R.; Araujo, R.; Mills, G. B.; Liotta, L. A.; Petricoin, E. F., III; Wulfskuhle, J. D. *Mol. Cell. Proteomics* **2005**, *4* (4), 346–55.
- Uhlen, M.; Ponten, F. *Mol. Cell. Proteomics* **2005**, *4* (4), 384–93.
- Utz, P. J. *Immunol. Rev.* **2005**, *204*, 264–82.
- Yuk, J. S.; Ha, K. S. *Exp. Mol. Med.* **2005**, *37* (1), 1–10.
- Nettikadan, S.; Radke, K.; Johnson, J.; Xu, J.; Lynch, M.; Mosher, C.; Henderson, E. *Mol. Cell. Proteomics* **2006**, *5* (5), 895–901.
- Collett, J. R.; Cho, E. J.; Ellington, A. D. *Methods* **2005**, *37* (1), 4–15.
- Jeong Cho, E.; Collett, J. R.; Szafranska, A. E.; Ellington, A. D. *Anal. Chim. Acta* **2006**, *564*, 82–90.
- Steinhauer, C.; Ressine, A.; Marko-Varga, G.; Laurell, T.; Borrebaeck, C. A.; Wingren, C. *Anal. Biochem.* **2005**, *341* (2), 204–13.
- Ptacek, J.; Devgan, G.; Michaud, G.; Zhu, H.; Zhu, X.; Fasolo, J.; Guo, H.; Jona, G.; Breitkreutz, A.; Sopko, R.; McCartney, R. R.; Schmidt, M. C.; Rachidi, N.; Lee, S. J.; Mah, A. S.; Meng, L.; Stark, M. J.; Stern, D. F.; De Virgilio, C.; Tyers, M.; Andrews, B.; Gerstein, M.; Schweitzer, B.; Predki, P. F.; Snyder, M. *Nature* **2005**, *438* (7068), 679–84.
- Hu, Q.; Noll, R. J.; Li, H.; Makarov, A.; Hardman, M.; Cooks, G. R. *J. Mass Spectrom.* **2005**, *40* (4), 430–43.
- Scigelova, M.; Makarov, A. *Proteomics* **2006**, *6* (Suppl 2), 16–21.
- Macek, B.; Waanders, L. F.; Olsen, J. V.; Mann, M. *Mol. Cell. Proteomics* **2006**, *5* (5), 949–58.
- Syka, J. E.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (26), 9528–33.
- Mikesh, L. M.; Ueberheide, B.; Chi, A.; Coon, J. J.; Syka, J. E.; Shabanowitz, J.; Hunt, D. F. *Biochim. Biophys. Acta* **2006**, *1764* (12), 1811–22.
- Liu, H.; Zhang, Y.; Wang, J.; Wang, D.; Zhou, C.; Cai, Y.; Qian, X. *Anal. Chem.* **2006**, *78* (18), 6614–21.
- Whetstone, P. A.; Butlin, N. G.; Corneille, T. M.; Meares, C. F. *Bioconjugate Chem.* **2004**, *15* (1), 3–6.
- Hoffert, J. D.; Pisitkun, T.; Wang, G.; Shen, R. F.; Knepper, M. A. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103* (18), 7159–64.
- Wang, G.; Wu, W. W.; Zeng, W.; Chou, C. L.; Shen, R. F. *J. Proteome Res.* **2006**, *5* (5), 1214–23.
- Johansson, C.; Samskog, J.; Sundstrom, L.; Wadensten, H.; Bjorkesten, L.; Flensburg, J. *Proteomics* **2006**, *6* (16), 4475–85.
- Pan, C.; Kora, G.; Tabb, D. L.; Pelletier, D. A.; McDonald, W. H.; Hurst, G. B.; Hettich, R. L.; Samatova, N. F. *Anal. Chem.* **2006**, *78* (20), 7110–20.
- Hattan, S. J.; Parker, K. C. *Anal. Chem.* **2006**, *78* (23), 7986–96.
- Hendrickson, E. L.; Xia, Q.; Wang, T.; Leigh, J. A.; Hackett, M. *Analyst* **2006**, *131* (12), 1335–41.
- Yan, W.; Chen, S. S. *Brief Funct. Genomic Proteomic* **2005**, *4* (1), 27–38.
- Mayya, V.; Rezaul, K.; Wu, L.; Fong, M. B.; Han, D. K. *Mol. Cell. Proteomics* **2006**, *5* (6), 1146–57.
- Kirkpatrick, D. S.; Gerber, S. A.; Gygi, S. P. *Methods* **2005**, *35* (3), 265–73.
- Kirkpatrick, D. S.; Hathaway, N. A.; Hanna, J.; Elsassner, S.; Rush, J.; Finley, D.; King, R. W.; Gygi, S. P. *Nat. Cell Biol.* **2006**, *8* (7), 700–10.
- Huang, S. Y.; Tsai, M. L.; Wu, C. J.; Hsu, J. L.; Ho, S. H.; Chen, S. H. *Proteomics* **2006**, *6* (6), 1722–34.
- Melanson, J. E.; Avery, S. L.; Pinto, D. M. *Proteomics* **2006**, *6* (16), 4466–74.
- Hsu, J. L.; Huang, S. Y.; Shiea, J. T.; Huang, W. Y.; Chen, S. H. *J. Proteome Res.* **2005**, *4* (1), 101–8.
- MacLellan, D. L.; Steen, H.; Adam, R. M.; Garlick, M.; Zurakowski, D.; Gygi, S. P.; Freeman, M. R.; Solomon, K. R. *Proteomics* **2005**, *5* (18), 4733–42.
- Moron, J. A.; Abul-Husn, N. S.; Rozenfeld, R.; Dolios, G.; Wang, R.; Devi, L. A. *Mol. Cell. Proteomics* **2007**, *6* (1), 29–42.
- Flory, M. R.; Lee, H.; Bonneau, R.; Mallick, P.; Serikawa, K.; Morris, D. R.; Aebersold, R. *Proteomics* **2006**, *6* (23), 6146–57.
- Aggarwal, K.; Choe, L. H.; Lee, K. H. *Brief Funct. Genomic Proteomic* **2006**, *5* (2), 112–20.
- Wiese, S.; Reidegeld, K. A.; Meyer, H. E.; Warscheid, B. *Proteomics* **2007**, *7* (3), 340–50.
- Hu, J.; Qian, J.; Borisov, O.; Pan, S.; Li, Y.; Liu, T.; Deng, L.; Wannemacher, K.; Kurnellas, M.; Patterson, C.; Elkabes, S.; Li, H. *Proteomics* **2006**, *6* (15), 4321–34.
- Liu, T.; D'Mello, V.; Deng, L.; Hu, J.; Ricardo, M.; Pan, S.; Lu, X.; Wadsworth, S.; Siekierka, J.; Birge, R.; Li, H. *J. Neurosci. Methods* **2006**, *158* (1), 22–9.
- Salim, K.; Kehoe, L.; Minkoff, M. S.; Bilslund, J. G.; Munoz-Sanjuán, I.; Guest, P. C. *Stem Cells Dev.* **2006**, *15* (3), 461–70.
- Sachon, E.; Mohammed, S.; Bache, N.; Jensen, O. N. *Rapid Commun. Mass Spectrom.* **2006**, *20* (7), 1127–34.
- Borner, G. H.; Harbour, M.; Hester, S.; Lilley, K. S.; Robinson, M. S. *J. Cell Biol.* **2006**, *175* (4), 571–8.
- Jagtap, P.; Michailidis, G.; Zielke, R.; Walker, A. K.; Patel, N.; Strahler, J. R.; Driks, A.; Andrews, P. C.; Maddock, J. R. *Proteomics* **2006**, *6* (19), 5199–211.
- Wu, W. W.; Wang, G.; Baek, S. J.; Shen, R. F. *J. Proteome Res.* **2006**, *5* (3), 651–8.
- Chong, P. K.; Gan, C. S.; Pham, T. K.; Wright, P. C. *J. Proteome Res.* **2006**, *5* (5), 1232–40.
- Molina, H.; Parmigiani, G.; Pandey, A. *Anal. Chem.* **2005**, *77* (9), 2739–44.
- Blagoev, B.; Mann, M. *Methods* **2006**, *40* (3), 243–50.
- Mann, M. *Nat. Rev. Mol. Cell Biol.* **2006**, *7* (12), 952–8.
- Everley, P. A.; Bakalarski, C. E.; Elias, J. E.; Waghorne, C. G.; Beausoleil, S. A.; Gerber, S. A.; Faherty, B. K.; Zetter, B. R.; Gygi, S. P. *J. Proteome Res.* **2006**, *5* (5), 1224–31.
- Gronborg, M.; Kristiansen, T. Z.; Iwahori, A.; Chang, R.; Reddy, R.; Sato, N.; Molina, H.; Jensen, O. N.; Hruban, R. H.; Goggins, M. G.; Maitra, A.; Pandey, A. *Mol. Cell. Proteomics* **2006**, *5* (1), 157–71.
- Hwang, S. I.; Lundgren, D. H.; Mayya, V.; Rezaul, K.; Cowan, A. E.; Eng, J. K.; Han, D. K. *Mol. Cell. Proteomics* **2006**, *5* (6), 1131–45.

- (72) Liang, X.; Zhao, J.; Hajivandi, M.; Wu, R.; Tao, J.; Amshey, J. W.; Pope, R. M. *J. Proteome Res.* **2006**, *5* (10), 2632–41.
- (73) Milner, E.; Barnea, E.; Beer, I.; Admon, A. *Mol. Cell. Proteomics* **2006**, *5* (2), 357–65.
- (74) Yan, Y.; Weaver, V. M.; Blair, I. A. *J. Proteome Res.* **2005**, *4* (6), 2007–14.
- (75) Amanchy, R.; Kalume, D. E.; Iwahori, A.; Zhong, J.; Pandey, A. *J. Proteome Res.* **2005**, *4* (5), 1661–71.
- (76) Grubler, A.; Olsen, J. V.; Mohammed, S.; Mortensen, P.; Faergeman, N. J.; Mann, M.; Jensen, O. N. *Mol. Cell. Proteomics* **2005**, *4* (3), 310–27.
- (77) Zhang, G.; Spellman, D. S.; Skolnik, E. Y.; Neubert, T. A. *J. Proteome Res.* **2006**, *5* (3), 581–8.
- (78) Doherty, M. K.; Whitehead, C.; McCormack, H.; Gaskell, S. J.; Beynon, R. J. *Proteomics* **2005**, *5* (2), 522–33.
- (79) Romijn, E. P.; Christis, C.; Wieffer, M.; Gouw, J. W.; Fullaondo, A.; van der Sluijs, P.; Braakman, I.; Heck, A. J. *Mol. Cell. Proteomics* **2005**, *4* (9), 1297–310.
- (80) Thiede, B.; Kretschmer, A.; Rudel, T. *Proteomics* **2006**, *6* (2), 614–22.
- (81) Neher, S. B.; Villen, J.; Oakes, E. C.; Bakalarski, C. E.; Sauer, R. T.; Gygi, S. P.; Baker, T. A. *Mol. Cell* **2006**, *22* (2), 193–204.
- (82) Grubler, A.; Schulze, W. X.; Matthiesen, R.; Mann, M.; Jensen, O. N. *Mol. Cell. Proteomics* **2005**, *4* (11), 1697–709.
- (83) Drukier, A. K.; Ossetrova, N.; Schors, E.; Brown, L. R.; Tomaszewski, J.; Sainsbury, R.; Godovac-Zimmermann, J. *J. Proteome Res.* **2005**, *4* (6), 2375–8.
- (84) Morelle, W.; Canis, K.; Chirat, F.; Faid, V.; Michalski, J. C. *Proteomics* **2006**, *6* (14), 3993–4015.
- (85) Prescher, J. A.; Bertozzi, C. R. *Cell* **2006**, *126* (5), 851–4.
- (86) Smith, J. C.; Figeys, D. *Mol. Biosyst.* **2006**, *2* (8), 364–70.
- (87) Wuhrer, M.; Deelder, A. M.; Hokke, C. H. *J. Chromatogr., B: Anal. Technol. Biomed. Life Sci.* **2005**, *825* (2), 124–33.
- (88) Yang, Z.; Hancock, W. S.; Chew, T. R.; Bonilla, L. *Proteomics* **2005**, *5* (13), 3353–66.
- (89) Qiu, R.; Regnier, F. E. *Anal. Chem.* **2005**, *77* (9), 2802–9.
- (90) Zhao, J.; Simeone, D. M.; Heidt, D.; Anderson, M. A.; Lubman, D. M. *J. Proteome Res.* **2006**, *5* (7), 1792–802.
- (91) Okanda, F. M.; El Rassi, Z. *Electrophoresis* **2006**, *27* (5–6), 1020–30.
- (92) Vosseller, K.; Trinidad, J. C.; Chalkley, R. J.; Specht, C. G.; Thalhammer, A.; Lynn, A. J.; Snedecor, J. H.; Guan, S.; Medzihradsky, K. F.; Maltby, D. A.; Schoepfer, R.; Burlingame, A. L. *Mol. Cell. Proteomics*. In press.
- (93) Zhang, H.; Li, X. J.; Martin, D. B.; Aebersold, R. *Nat. Biotechnol.* **2003**, *21* (6), 660–6.
- (94) Zhang, H.; Yi, E. C.; Li, X. J.; Mallick, P.; Kelly-Spratt, K. S.; Masselon, C. D.; Camp, D. G., II; Smith, R. D.; Kemp, C. J.; Aebersold, R. *Mol. Cell. Proteomics* **2005**, *4* (2), 144–55.
- (95) Sun, B.; Ranish, J. A.; Utleg, A. G.; White, J. T.; Yan, X.; Lin, B.; Hood, L. *Mol. Cell. Proteomics* **2007**, *6* (1), 141–9.
- (96) Ramachandran, P.; Boonthuep, P.; Xie, Y.; Sondej, M.; Wong, D. T.; Loo, J. A. *J. Proteome Res.* **2006**, *5* (6), 1493–503.
- (97) Sprung, R.; Nandi, A.; Chen, Y.; Kim, S. C.; Barma, D.; Falck, J. R.; Zhao, Y. *J. Proteome Res.* **2005**, *4* (3), 950–7.
- (98) Chan, K. S.; Kho, Y.; Barma, D.; Falck, J.; Zhao, Y. *Methods Enzymol.* **2005**, *407*, 629–37.
- (99) Nally, J. E.; Whitelegge, J. P.; Aguilera, R.; Pereira, M. M.; Blanco, D. R.; Lovett, M. A. *Proteomics* **2005**, *5* (1), 144–52.
- (100) Hoffman, M. D.; Kast, J. *Mass Spectrom.* **2006**, *41* (2), 229–41.
- (101) Drisdell, R. C.; Alexander, J. K.; Sayeed, A.; Green, W. N. *Methods* **2006**, *40* (2), 127–34.
- (102) Roth, A. F.; Wan, J.; Green, W. N.; Yates, J. R.; Davis, N. G. *Methods* **2006**, *40* (2), 135–42.
- (103) Beck, H. C.; Nielsen, E. C.; Matthiesen, R.; Jensen, L. H.; Sehested, M.; Finn, P.; Grauslund, M.; Hansen, A. M.; Jensen, O. N. *Mol. Cell. Proteomics* **2006**, *5* (7), 1314–25.
- (104) Falb, M.; Aivaliotis, M.; Garcia-Rizo, C.; Bisle, B.; Tebbe, A.; Klein, C.; Konstantinidis, K.; Siedler, F.; Pfeiffer, F.; Oesterheld, D. *J. Mol. Biol.* **2006**, *362* (5), 915–24.
- (105) Lee, Y. J.; Rice, R. H.; Lee, Y. M. *Mol. Cell. Proteomics* **2006**, *5* (5), 789–800.
- (106) Balkhi, M. Y.; Trivedi, A. K.; Geletu, M.; Christopheit, M.; Bohlander, S. K.; Behre, H. M.; Behre, G. *Oncogene* **2006**, *25* (53), 7041–58.
- (107) Kim, S. C.; Sprung, R.; Chen, Y.; Xu, Y.; Ball, H.; Pei, J.; Cheng, T.; Kho, Y.; Xiao, H.; Xiao, L.; Grishin, N. V.; White, M.; Yang, X. J.; Zhao, Y. *Mol. Cell. Proteomics* **2006**, *23* (4), 607–18.
- (108) Iwabata, H.; Yoshida, M.; Komatsu, Y. *Proteomics* **2005**, *5* (18), 4653–64.
- (109) Ciechanover, A. *Proc. Am. Thorac. Soc.* **2006**, *3* (1), 21–31.
- (110) Xu, P.; Peng, J. *Biochim. Biophys. Acta* **2006**, *1764* (12), 1940–7.
- (111) Tagwerker, C.; Flick, K.; Cui, M.; Guerrero, C.; Dou, Y.; Auer, B.; Baldi, P.; Huang, L.; Kaiser, P. *Mol. Cell. Proteomics* **2006**, *5* (4), 737–48.
- (112) Kirkpatrick, D. S.; Weldon, S. F.; Tsapralis, G.; Liebler, D. C.; Gandolfi, A. *J. Proteomics* **2005**, *5* (8), 2104–11.
- (113) Vertegaal, A. C.; Andersen, J. S.; Ogg, S. C.; Hay, R. T.; Mann, M.; Lamond, A. I. *Mol. Cell. Proteomics* **2006**, *5* (12), 2298–310.
- (114) Denison, C.; Rudner, A. D.; Gerber, S. A.; Bakalarski, C. E.; Moazed, D.; Gygi, S. P. *Mol. Cell. Proteomics* **2005**, *4* (3), 246–54.
- (115) Rosas-Acosta, G.; Russell, W. K.; Deyrieux, A.; Russell, D. H.; Wilson, V. G. *Mol. Cell. Proteomics* **2005**, *4* (1), 56–72.
- (116) Vasilescu, J.; Smith, J. C.; Ethier, M.; Figeys, D. *J. Proteome Res.* **2005**, *4* (6), 2192–200.
- (117) Matsumoto, M.; Hatakeyama, S.; Oyamada, K.; Oda, Y.; Nishimura, T.; Nakayama, K. I. *Proteomics* **2005**, *5* (16), 4145–51.
- (118) Wang, J.; Zhang, Y.; Jiang, H.; Cai, Y.; Qian, X. *Proteomics* **2006**, *6* (2), 404–11.
- (119) Feuerstein, I.; Morandell, S.; Stecher, G.; Huck, C. W.; Stasyk, T.; Huang, H. L.; Huber, L. A.; Bonn, G. K. *Proteomics* **2005**, *5* (1), 46–54.
- (120) Feng, S.; Pan, C.; Jiang, X.; Xu, S.; Zhou, H.; Ye, M.; Zou, H. *Proteomics* **2007**, *7* (3), 351–60.
- (121) Aprilita, N. H.; Huck, C. W.; Bakry, R.; Feuerstein, I.; Stecher, G.; Morandell, S.; Huang, H. L.; Stasyk, T.; Huber, L. A.; Bonn, G. K. *J. Proteome Res.* **2005**, *4* (6), 2312–9.
- (122) Barnouin, K. N.; Hart, S. R.; Thompson, A. J.; Okuyama, M.; Waterfield, M.; Cramer, R. *Proteomics* **2005**, *5* (17), 4376–88.
- (123) Guerrero, I. C.; Predic-Atkinson, J.; Kleiner, O.; Soskic, V.; Godovac-Zimmermann, J. *J. Proteome Res.* **2005**, *4* (5), 1545–53.
- (124) Wolschin, F.; Wienkoop, S.; Weckwerth, W. *Proteomics* **2005**, *5* (17), 4389–97.
- (125) Arrigoni, G.; Resjo, S.; Levander, F.; Nilsson, R.; Degerman, E.; Quadroni, M.; Pinna, L. A.; James, P. *Proteomics* **2006**, *6* (3), 757–66.
- (126) Amoresano, A.; Monti, G.; Cirulli, C.; Marino, G. *Rapid Commun. Mass Spectrom.* **2006**, *20* (9), 1400–4.
- (127) Unwin, R. D.; Griffiths, J. R.; Leverenz, M. K.; Grallert, A.; Hagan, I. M.; Whetton, A. D. *Mol. Cell. Proteomics* **2005**, *4* (8), 1134–44.
- (128) Gevaert, K.; Staes, A.; Van Damme, J.; De Groot, S.; Hugelier, K.; Demol, H.; Martens, L.; Goethals, M.; Vandekerckhove, J. *Proteomics* **2005**, *5* (14), 3589–99.
- (129) Beausoleil, S. A.; Villen, J.; Gerber, S. A.; Rush, J.; Gygi, S. P. *Nat. Biotechnol.* **2006**, *24* (10), 1285–92.
- (130) Schwartz, D.; Gygi, S. P. *Nat. Biotechnol.* **2005**, *23* (11), 1391–8.
- (131) Wang, J.; Rao, S.; Chu, J.; Shen, X.; Levasseur, D. N.; Theunissen, T. W.; Orkin, S. H. *Nature* **2006**, *444* (7117), 364–8.
- (132) Takahashi, K.; Yamanaka, S. *Cell* **2006**, *126* (4), 663–76.
- (133) Ewing, R. M.; Chu, P.; Elisma, F.; Li, H.; Taylor, P.; Climie, S.; McBroom-Cerajewski, L.; Robinson, M. D.; O'Connor, L.; Li, M.; Taylor, R.; Dharsee, M.; Ho, Y.; Heilbut, A.; Moore, L.; Zhang, S.; Ornatsky, O.; Bukhman, Y. V.; Ethier, M.; Sheng, Y.; Vasilescu, J.; Abu-Farha, M.; Lambert, J. P.; Duewel, H. S.; Stewart, II; Kuehl, B.; Hogue, K.; Colwill, K.; Gladwish, K.; Muskat, B.; Kinach, R.; Adams, S. L.; Moran, M. F.; Morin, G. B.; Topaloglu, T.; Figeys, D. *Mol. Syst. Biol.* **2007**, *3*, 89.
- (134) Barrios-Rodiles, M.; Brown, K. R.; Ozdamar, B.; Bose, R.; Liu, Z.; Donovan, R. S.; Shinjo, F.; Liu, Y.; Dombow, J.; Taylor, I. W.; Luga, V.; Przulj, N.; Robinson, M.; Suzuki, H.; Hayashizaki, Y.; Jurisica, I.; Wrana, J. L. *Science* **2005**, *307* (5715), 1621–5.
- (135) Rual, J. F.; Venkatesan, K.; Hao, T.; Hirozane-Kishikawa, T.; Dricot, A.; Li, N.; Berriz, G. F.; Gibbons, F. D.; Dreze, M.; Ayivi-Guedehoussou, N.; Klitborg, N.; Simon, C.; Boxem, M.; Milstein, S.; Rosenberg, J.; Goldberg, D. S.; Zhang, L. V.; Wong, S. L.; Franklin, G.; Li, S.; Albaladejo, J. S.; Lim, J.; Fraughton, C.; Llamosas, E.; Cevik, S.; Bex, C.; Lamesch, P.; Sikorski, R. S.; Vandenhaute, J.; Zoghbi, H. Y.; Smolyar, A.; Bosak, S.; Sequerra, R.; Doucette-Stamm, L.; Cusick, M. E.; Hill, D. E.; Roth, F. P.; Vidal, M. *Nature* **2005**, *437* (7062), 1173–8.
- (136) Stelzl, U.; Worm, U.; Lalowski, M.; Haenig, C.; Brembeck, F. H.; Goehler, H.; Stroedicke, M.; Zenkner, M.; Schoenherr, A.; Koepfle, S.; Timm, J.; Mintzlaff, S.; Abraham, C.; Bock, N.; Kietzmann, S.; Goedde, A.; Toksoz, E.; Droegge, A.; Krobitsch, S.; Korn, B.; Birchmeier, W.; Lehrach, H.; Wanker, E. E. *Cell* **2005**, *122* (6), 957–68.
- (137) Lim, J.; Hao, T.; Shaw, C.; Patel, A. J.; Szabo, G.; Rual, J. F.; Fisk, C. J.; Li, N.; Smolyar, A.; Hill, D. E.; Barabasi, A. L.; Vidal, M.; Zoghbi, H. Y. *Cell* **2006**, *125* (4), 801–14.
- (138) Selbach, M.; Mann, M. *Nat. Methods* **2006**, *3* (12), 981–3.
- (139) Krogan, N. J.; Cagney, G.; Yu, H.; Zhong, G.; Guo, X.; Ignatchenko, A.; Li, J.; Pu, S.; Datta, N.; Tikuisis, A. P.; Punna, T.; Peregrin-Alvarez, J. M.; Shales, M.; Zhang, X.; Davey, M.; Robinson, M. D.; Paccanaro, A.; Bray, J. E.; Sheung, A.; Beattie, B.; Richards, D. P.; Canadien, V.; Lalev, A.; Mena, F.; Wong, P.; Starostine, A.; Canete, M. M.; Vlasblom, J.; Wu, S.; Orsi, C.; Collins, S. R.; Chandran, S.; Haw, R.; Rilstone, J. J.; Gandi, K.; Thompson, N. J.; Musso, G.; St. Onge, P.; Ghanny, S.; Lam, M. H.; Butland, G.; Altaf-Ul, A. M.; Kanaya, S.; Shilatifard, A.; O'Shea, E.; Weissman, J. S.; Ingles, C. J.; Hughes, T. R.; Parkinson, J.; Gerstein, M.; Wodak, S. J.; Emili, A.; Greenblatt, J. F. *Nature* **2006**, *440* (7084), 637–43.
- (140) Gavin, A. C.; Aloy, P.; Grandi, P.; Krause, R.; Boesche, M.; Marzioch, M.; Rau, C.; Jensen, L. J.; Bastuck, S.; Dumppelfeld, B.; Edelmann, A.; Heurter, M. A.; Hoffman, V.; Hoefert, C.; Klein, K.; Hudak, M.; Michon, A. M.; Schelder, M.; Schirle, M.; Remor, M.; Rudi, T.; Hooper, S.; Bauer, A.; Bouwmeester, T.; Casari, G.; Drewes, G.; Neubauer, G.; Rick, J. M.; Kuster, B.;

- Bork, P.; Russell, R. B.; Superti-Furga, G. *Nature* **2006**, *440* (7084), 631–6.
- (141) Brohee, S.; van Helden, J. *BMC Bioinf.* **2006**, *7*, 488.
- (142) Collins, S. R.; Kemmeren, P.; Zhao, X. C.; Greenblatt, J. F.; Spencer, F.; Holstege, F. C.; Weissman, J. S.; Krogan, N. J. *Mol. Cell. Proteomics* **2007**, *6* (3), 439–50.
- (143) Fleischer, T. C.; Weaver, C. M.; McAfee, K. J.; Jennings, J. L.; Link, A. J. *Genes Dev.* **2006**, *20* (10), 1294–307.
- (144) Uetz, P.; Dong, Y. A.; Zeretzke, C.; Atzler, C.; Baiker, A.; Berger, B.; Rajagopala, S. V.; Roupelieva, M.; Rose, D.; Fossum, E.; Haas, J. *Science* **2006**, *311* (5758), 239–42.
- (145) Mayer, D.; Molawi, K.; Martinez-Sobrido, L.; Ghanem, A.; Thomas, S.; Baginsky, S.; Grossmann, J.; Garcia-Sastre, A.; Schwemmler, M. *J. Proteome Res.* **2007**, *6* (2), 672–82.
- (146) Butland, G.; Peregrin-Alvarez, J. M.; Li, J.; Yang, W.; Yang, X.; Canadien, V.; Starostine, A.; Richards, D.; Beattie, B.; Krogan, N.; Davey, M.; Parkinson, J.; Greenblatt, J.; Emili, A. *Nature* **2005**, *433* (7025), 531–7.
- (147) Arifuzzaman, M.; Maeda, M.; Itoh, A.; Nishikata, K.; Takita, C.; Saito, R.; Ara, T.; Nakahigashi, K.; Huang, H. C.; Hirai, A.; Tsuzuki, K.; Nakamura, S.; Altaf-Ul-Amin, M.; Oshima, T.; Baba, T.; Yamamoto, N.; Kawamura, T.; Ioka-Nakamichi, T.; Kitagawa, M.; Tomita, M.; Kanaya, S.; Wada, C.; Mori, H. *Genome Res.* **2006**, *16* (5), 686–91.
- (148) Riedel, C. G.; Katis, V. L.; Katou, Y.; Mori, S.; Itoh, T.; Helmhart, W.; Galova, M.; Petronczki, M.; Gregan, J.; Cetin, B.; Mudrak, I.; Ogris, E.; Mechtler, K.; Pelletier, L.; Buchholz, F.; Shirahige, K.; Nasmyth, K. *Nature* **2006**, *441* (7089), 53–61.
- (149) Denis, G. V.; McComb, M. E.; Faller, D. V.; Sinha, A.; Romesser, P. B.; Costello, C. E. *J. Proteome Res.* **2006**, *5* (3), 502–11.
- (150) Khorasanizadeh, S. *Cell* **2004**, *116* (2), 259–72.
- (151) Bernstein, E.; Hake, S. B. *Biochem. Cell Biol.* **2006**, *84* (4), 505–17.
- (152) Foltz, D. R.; Jansen, L. E.; Black, B. E.; Bailey, A. O.; Yates, J. R., III; Cleveland, D. W. *Nat. Cell Biol.* **2006**, *8* (5), 458–69.
- (153) Du, Y. C.; Gu, S.; Zhou, J.; Wang, T.; Cai, H.; Macinnes, M. A.; Bradbury, E. M.; Chen, X. *Mol. Cell. Proteomics* **2006**, *5* (6), 1033–44.
- (154) Dirksen, E. H.; Pinkse, M. W.; Rijkers, D. T.; Cloos, J.; Liskamp, R. M.; Slijper, M.; Heck, A. J. *J. Proteome Res.* **2006**, *5* (9), 2380–8.
- (155) Blais, A.; Dynlacht, B. D. *Genes Dev.* **2005**, *19* (13), 1499–511.
- (156) Kim, J.; Bhinge, A. A.; Morgan, X. C.; Iyer, V. R. *Nat. Methods* **2005**, *2* (1), 47–53.
- (157) Wei, C. L.; Wu, Q.; Vega, V. B.; Chiu, K. P.; Ng, P.; Zhang, T.; Shahab, A.; Yong, H. C.; Fu, Y.; Weng, Z.; Liu, J.; Zhao, X. D.; Chew, J. L.; Lee, Y. L.; Kuznetsov, V. A.; Sung, W. K.; Miller, L. D.; Lim, B.; Liu, E. T.; Yu, Q.; Ng, H. H.; Ruan, Y. *Cell* **2006**, *124* (1), 207–19.
- (158) Zeller, K. I.; Zhao, X.; Lee, C. W.; Chiu, K. P.; Yao, F.; Yustein, J. T.; Ooi, H. S.; Orlov, Y. L.; Shahab, A.; Yong, H. C.; Fu, Y.; Weng, Z.; Kuznetsov, V. A.; Sung, W. K.; Ruan, Y.; Dang, C. V.; Wei, C. L. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103* (47), 17834–9.
- (159) Loh, Y. H.; Wu, Q.; Chew, J. L.; Vega, V. B.; Zhang, W.; Chen, X.; Bourque, G.; George, J.; Leong, B.; Liu, J.; Wong, K. Y.; Sung, K. W.; Lee, C. W.; Zhao, X. D.; Chiu, K. P.; Lipovich, L.; Kuznetsov, V. A.; Robson, P.; Stanton, L. W.; Wei, C. L.; Ruan, Y.; Lim, B.; Ng, H. H. *Nat. Genet.* **2006**, *38* (4), 431–40.
- (160) Ho, S. W.; Jona, G.; Chen, C. T.; Johnston, M.; Snyder, M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103* (26), 9940–5.
- (161) Mukherjee, S.; Berger, M. F.; Jona, G.; Wang, X. S.; Muzzey, D.; Snyder, M.; Young, R. A.; Bulky, M. L. *Nat. Genet.* **2004**, *36* (12), 1331–9.
- (162) Berger, M. F.; Philippakis, A. A.; Qureshi, A. M.; He, F. S.; Estep, P. W., III; Bulky, M. L. *Nat. Biotechnol.* **2006**, *24* (11), 1429–35.
- (163) Godtl, K.; Gruss, O. J.; Eickhoff, J.; Wissing, J.; Blencke, S.; Weber, M.; Degen, H.; Brehmer, D.; Orfi, L.; Horvath, Z.; Keri, G.; Muller, S.; Cotten, M.; Ullrich, A.; Daub, H. *Cancer Res.* **2005**, *65* (15), 6919–26.
- (164) Kosaka, T.; Okuyama, R.; Sun, W.; Ogata, T.; Harada, J.; Araki, K.; Izumi, M.; Yoshida, T.; Okuno, A.; Fujiwara, T.; Ohsumi, J.; Ichikawa, K. *Anal. Chem.* **2005**, *77* (7), 2050–5.
- (165) Salisbury, C. M.; Cravatt, B. F. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104* (4), 1171–6.
- (166) Fabian, M. A.; Biggs, W. H., III; Treiber, D. K.; Atteridge, C. E.; Azimioara, M. D.; Benedetti, M. G.; Carter, T. A.; Ciceri, P.; Edeen, P. T.; Floyd, M.; Ford, J. M.; Galvin, M.; Gerlach, J. L.; Grotzfeld, R. M.; Herrgard, S.; Insko, D. E.; Insko, M. A.; Lai, A. G.; Lelias, J. M.; Mehta, S. A.; Milanov, Z. V.; Velasco, A. M.; Wodicka, L. M.; Patel, H. K.; Zarrinkar, P. P.; Lockhart, D. J. *Nat. Biotechnol.* **2005**, *23* (3), 329–36.
- (167) Sieber, S. A.; Niessen, S.; Hoover, H. S.; Cravatt, B. F. *Nat. Chem. Biol.* **2006**, *2* (5), 274–81.
- (168) Kumar, S.; Zhou, B.; Liang, F.; Yang, H.; Wang, W. Q.; Zhang, Z. Y. *J. Proteome Res.* **2006**, *5* (8), 1898–905.
- (169) Suchanek, M.; Radzikowska, A.; Thiele, C. *Nat. Methods* **2005**, *2* (4), 261–7.
- (170) Liu, W.; Brock, A.; Chen, S.; Chen, S.; Schultz, P. G. *Nat. Methods* **2007**, *4* (3), 239–44.
- (171) Di Ventura, B.; Lemerle, C.; Michalodimitrakis, K.; Serrano, L. *Nature* **2006**, *443* (7111), 527–33.
- (172) Kitano, H. *Science* **2002**, *295* (5560), 1662–4.
- (173) Belle, A.; Tanay, A.; Bitincka, L.; Shamir, R.; O’Shea, E. K. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103* (35), 13004–9.
- (174) Sopko, R.; Huang, D.; Preston, N.; Chua, G.; Papp, B.; Kafadar, K.; Snyder, M.; Oliver, S. G.; Cyert, M.; Hughes, T. R.; Boone, C.; Andrews, B. *Mol. Cell* **2006**, *21* (3), 319–30.
- (175) Jones, R. B.; Gordus, A.; Krall, J. A.; MacBeath, G. *Nature* **2006**, *439* (7073), 168–74.
- (176) Chu, D. S.; Liu, H.; Nix, P.; Wu, T. F.; Ralston, E. J.; Yates, J. R., III; Meyer, B. J. *Nature* **2006**, *443* (7107), 101–5.
- (177) Stuart, L. M.; Boulais, J.; Charriere, G. M.; Hennessy, E. J.; Brunet, S.; Jutras, I.; Goyette, G.; Rondeau, C.; Letarte, S.; Huang, H.; Ye, P.; Morales, F.; Kocks, C.; Bader, J. S.; Desjardins, M.; Ezekowitz, R. A. *Nature* **2007**, *445* (7123), 95–101.
- (178) Reichert, J. M. *Nat. Rev. Drug Discovery* **2003**, *2* (9), 695–702.
- (179) West-Nielsen, M.; Hogdall, E. V.; Marchiori, E.; Hogdall, C. K.; Schou, C.; Heegaard, N. H. *Anal. Chem.* **2005**, *77* (16), 5114–23.
- (180) Hsieh, S. Y.; Chen, R. K.; Pan, Y. H.; Lee, H. L. *Proteomics* **2006**, *6* (10), 3189–98.
- (181) Rai, A. J.; Gelfand, C. A.; Haywood, B. C.; Warunek, D. J.; Yi, J.; Schuchard, M. D.; Mehig, R. J.; Cockrill, S. L.; Scott, G. B.; Tammen, H.; Schulz-Knappe, P.; Speicher, D. W.; Vitzthum, F.; Haab, B. B.; Siest, G.; Chan, D. W. *Proteomics* **2005**, *5* (13), 3262–77.
- (182) Villanueva, J.; Philip, J.; Chaparro, C. A.; Li, Y.; Toledo-Crow, R.; DeNoyer, L.; Fleisher, M.; Robbins, R. J.; Tempst, P. *J. Proteome Res.* **2005**, *4* (4), 1060–72.
- (183) Anderson, N. L.; Anderson, N. G. *Mol. Cell. Proteomics* **2002**, *1* (11), 845–67.
- (184) Colantonio, D. A.; Dunkinson, C.; Bovenkamp, D. E.; Van Eyk, J. E. *Proteomics* **2005**, *5* (15), 3831–5.
- (185) Chen, Y. Y.; Lin, S. Y.; Yeh, Y. Y.; Hsiao, H. H.; Wu, C. Y.; Chen, S. T.; Wang, A. H. *Electrophoresis* **2005**, *26* (11), 2117–27.
- (186) Pieper, R.; Su, Q.; Gatlin, C. L.; Huang, S. T.; Anderson, N. L.; Steiner, S. *Proteomics* **2003**, *3* (4), 422–32.
- (187) Zolotarova, N.; Martosella, J.; Nicol, G.; Bailey, J.; Boyes, B. E.; Barrett, W. C. *Proteomics* **2005**, *5* (13), 3304–13.
- (188) Travis, J.; Pannell, R. *Clin. Chim. Acta* **1973**, *49* (1), 49–52.
- (189) Thresher, W. C.; Swaisgood, H. E. *J. Mol. Recognit.* **1990**, *3* (5–6), 220–8.
- (190) Bjorhull, K.; Miliotis, T.; Davidsson, P. *Proteomics* **2005**, *5* (1), 307–17.
- (191) Echan, L. A.; Tang, H. Y.; Ali-Khan, N.; Lee, K.; Speicher, D. W. *Proteomics* **2005**, *5* (13), 3292–303.
- (192) Yocum, A. K.; Yu, K.; Oe, T.; Blair, I. A. *J. Proteome Res.* **2005**, *4* (5), 1722–31.
- (193) Huang, L.; Harvie, G.; Feitelson, J. S.; Gramatikoff, K.; Herold, D. A.; Allen, D. L.; Amunngama, R.; Hagler, R. A.; Pisano, M. R.; Zhang, W. W.; Fang, X. *Proteomics* **2005**, *5* (13), 3314–28.
- (194) Liu, T.; Qian, W. J.; Mottaz, H. M.; Gritsenko, M. A.; Norbeck, A. D.; Moore, R. J.; Purvine, S. O.; Camp, D. G., II; Smith, R. D. *Mol. Cell. Proteomics* **2006**, *5* (11), 2167–74.
- (195) Gong, Y.; Li, X.; Yang, B.; Ying, W.; Li, D.; Zhang, Y.; Dai, S.; Cai, Y.; Wang, J.; He, F.; Qian, X. *J. Proteome Res.* **2006**, *5* (6), 1379–87.
- (196) Duan, X.; Yarmush, D.; Berthiaume, F.; Jayaraman, A.; Yarmush, M. L. *Proteomics* **2005**, *5* (15), 3991–4000.
- (197) Freeman, W. M.; Lull, M. E.; Guilford, M. T.; Vrana, K. E. *Proteomics* **2006**, *6* (10), 3109–13.
- (198) Whiteaker, J. R.; Zhang, H.; Eng, J. K.; Fang, R.; Piening, B. D.; Feng, L. C.; Lorentzen, T. D.; Schoenherr, R. M.; Keane, J. F.; Holzman, T.; Fitzgibbon, M.; Lin, C.; Zhang, H.; Cooke, K.; Liu, T.; Camp, D. G., II; Anderson, L.; Watts, J.; Smith, R. D.; McIntosh, M. W.; Paulovich, A. G. *J. Proteome Res.* **2007**, *6* (2), 828–36.
- (199) Liu, T.; Qian, W. J.; Strittmatter, E. F.; Camp, D. G., II; Anderson, G. A.; Thrall, B. D.; Smith, R. D. *Anal. Chem.* **2004**, *76* (18), 5345–53.
- (200) Baumann, S.; Ceglarek, U.; Fiedler, G. M.; Lembcke, J.; Leichtle, A.; Thiery, J. *Clin. Chem.* **2005**, *51* (6), 973–80.
- (201) Yin, H.; Killeen, K.; Brennen, R.; Sobek, D.; Werlich, M.; van de Goor, T. *Anal. Chem.* **2005**, *77* (2), 527–33.
- (202) Hardouin, J.; Duchateau, M.; Joubert-Caron, R.; Caron, M. *Rapid Commun. Mass Spectrom.* **2006**, *20* (21), 3236–44.
- (203) Fortier, M. H.; Bonnell, E.; Goodley, P.; Thibault, P. *Anal. Chem.* **2005**, *77* (6), 1631–40.
- (204) Lazar, I. M.; Trisiripisal, P.; Sarvaiya, H. A. *Anal. Chem.* **2006**, *78* (15), 5513–24.
- (205) Herr, A. E.; Hatch, A. V.; Throckmorton, D. J.; Tran, H. M.; Brennan, J. S.; Giannobile, W. V.; Singh, A. K. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104* (13), 5268–73.
- (206) Yang, S.; Undar, A.; Zahn, J. D. *Lab Chip* **2006**, *6* (7), 871–80.
- (207) Zeiler, K.; Pascher, G.; Hannig, K. *Hoppe Seylers Z. Physiol. Chem.* **1970**, *351* (4), 435–47.
- (208) Hoffmann, P.; Ji, H.; Moritz, R. L.; Connolly, L. M.; Frecklington, D. F.; Layton, M. J.; Eddes, J. S.; Simpson, R. J. *Proteomics* **2001**, *1* (7), 807–18.
- (209) Moritz, R. L.; Clippingdale, A. B.; Kapp, E. A.; Eddes, J. S.; Ji, H.; Gilbert, S.; Connolly, L. M.; Simpson, R. J. *Proteomics* **2005**, *5* (13), 3402–13.

- (210) Moritz, R. L.; Ji, H.; Schutz, F.; Connolly, L. M.; Kapp, E. A.; Speed, T. P.; Simpson, R. J. *Anal. Chem.* **2004**, *76* (16), 4811–24.
- (211) Sanchez-Carbayo, M.; Socci, N. D.; Lozano, J. J.; Haab, B. B.; Cordon-Cardo, C. *Am. J. Pathol.* **2006**, *168* (1), 93–103.
- (212) Shafer, M. W.; Mangold, L.; Partin, A. W.; Haab, B. B. *Prostate* **2007**, *67* (3), 255–67.
- (213) Horn, S.; Lueking, A.; Murphy, D.; Staudt, A.; Gutjahr, C.; Schulte, K.; Konig, A.; Landsberger, M.; Lehrach, H.; Felix, S. B.; Cahill, D. J. *Proteomics* **2006**, *6* (2), 605–13.
- (214) Andresen, H.; Grotzinger, C.; Zarse, K.; Kreuzer, O. J.; Ehrenreich-Forster, E.; Bier, F. F. *Proteomics* **2006**, *6* (5), 1376–84.
- (215) Gannot, G.; Tangrea, M. A.; Gillespie, J. W.; Erickson, H. S.; Wallis, B. S.; Leakan, R. A.; Knezevic, V.; Hartmann, D. P.; Chuauqui, R. F.; Emmert-Buck, M. R. *J. Mol. Diagn.* **2005**, *7* (4), 427–36.
- (216) Villanueva, J.; Shaffer, D. R.; Philip, J.; Chaparro, C. A.; Erdjument-Bromage, H.; Olshen, A. B.; Fleisher, M.; Lilja, H.; Brogi, E.; Boyd, J.; Sanchez-Carbayo, M.; Holland, E. C.; Cordon-Cardo, C.; Scher, H. I.; Tempst, P. *J. Clin. Invest.* **2006**, *116* (1), 271–84.
- (217) Diamandis, E. P. *J. Proteome Res.* **2006**, *5* (9), 2079–82.
- (218) Robbins, R. J.; Villanueva, J.; Tempst, P. *J. Clin. Oncol.* **2005**, *23* (22), 4835–7.
- (219) Diamandis, E. P. *Mol. Cell. Proteomics* **2004**, *3* (4), 367–78.
- (220) Diamandis, E. P. *J. Natl. Cancer Inst.* **2004**, *96* (5), 353–6.
- (221) Rai, A. J.; Stemmer, P. M.; Zhang, Z.; Adam, B. L.; Morgan, W. T.; Caffrey, R. E.; Podust, V. N.; Patel, M.; Lim, L. Y.; Shipulina, N. V.; Chan, D. W.; Semmes, O. J.; Leung, H. C. *Proteomics* **2005**, *5* (13), 3467–74.
- (222) Stajich, J. E.; Lapp, H. *Brief Bioinf.* **2006**, *7* (3), 287–96.
- (223) Field, D.; Tiwari, B.; Booth, T.; Houten, S.; Swan, D.; Bertrand, N.; Thurston, M. *Nat. Biotechnol.* **2006**, *24* (7), 801–3.
- (224) Spjuth, O.; Helmus, T.; Willighagen, E. L.; Kuhn, S.; Eklund, M.; Wagener, J.; Murray-Rust, P.; Steinbeck, C.; Wikberg, J. E. *BMC Bioinf.* **2007**, *8*, 59.
- (225) Saito, A.; Nagasaki, M.; Oyama, M.; Kozuka-Hata, H.; Semba, K.; Sugano, S.; Yamamoto, T.; Miyano, S. *BMC Bioinf.* **2007**, *8*, 15.
- (226) Katajamaa, M.; Oresic, M. *BMC Bioinf.* **2005**, *6*, 179.
- (227) Schulze, W. X.; Mann, M. *J. Biol. Chem.* **2004**, *279* (11), 10756–64.
- (228) Ross, P. L.; Huang, Y. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Purkayastha, S.; Juhasz, P.; Martin, S.; Bartlet-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J. *Mol. Cell. Proteomics* **2004**, *3* (12), 1154–69.
- (229) Shadforth, I. P.; Dunkley, T. P.; Lilley, K. S.; Bessant, C. *BMC Genomics* **2005**, *6*, 145.
- (230) Lin, W. T.; Hung, W. N.; Yian, Y. H.; Wu, K. P.; Han, C. L.; Chen, Y. R.; Chen, Y. J.; Sung, T. Y.; Hsu, W. L. *J. Proteome Res.* **2006**, *5* (9), 2328–38.
- (231) Garden, P.; Alm, R.; Hakkinen, J. *Bioinformatics* **2005**, *21* (9), 2085–7.
- (232) Taylor, C. F.; Paton, N. W.; Garwood, K. L.; Kirby, P. D.; Stead, D. A.; Yin, Z.; Deutsch, E. W.; Selway, L.; Walker, J.; Riba-Garcia, I.; Mohammed, S.; Deery, M. J.; Howard, J. A.; Dunkley, T.; Aebersold, R.; Kell, D. B.; Lilley, K. S.; Roepstorff, P.; Yates, J. R., III; Brass, A.; Brown, A. J.; Cash, P.; Gaskell, S. J.; Hubbard, S. J.; Oliver, S. G. *Nat. Biotechnol.* **2003**, *21* (3), 247–54.
- (233) Rauch, A.; Bellew, M.; Eng, J.; Fitzgibbon, M.; Holzman, T.; Hussey, P.; Igra, M.; Maclean, B.; Lin, C. W.; Detter, A.; Fang, R.; Faca, V.; Gafken, P.; Zhang, H.; Whiteaker, J.; States, D.; Hanash, S.; Paulovich, A.; McIntosh, M. W. *J. Proteome Res.* **2006**, *5* (1), 112–21.
- (234) Cerami, E. G.; Bader, G. D.; Gross, B. E.; Sander, C. *BMC Bioinf.* **2006**, *7*, 497.
- (235) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. *Genome Res.* **2003**, *13* (11), 2498–504.
- (236) Hermjakob, H.; Montecchi-Palazzi, L.; Bader, G.; Wojcik, J.; Salwinski, L.; Ceol, A.; Moore, S.; Orchard, S.; Sarkans, U.; von Mering, C.; Roechert, B.; Poux, S.; Jung, E.; Mersch, H.; Kersey, P.; Lappe, M.; Li, Y.; Zeng, R.; Rana, D.; Nikolski, M.; Husi, H.; Brun, C.; Shanker, K.; Grant, S. G.; Sander, C.; Bork, P.; Zhu, W.; Pandey, A.; Brazma, A.; Jacq, B.; Vidal, M.; Sherman, D.; Legrain, P.; Cesareni, G.; Xenarios, I.; Eisenberg, D.; Steipe, B.; Hogue, C.; Apweiler, R. *Nat. Biotechnol.* **2004**, *22* (2), 177–83.
- (237) Stanislaus, R.; Chen, C.; Franklin, J.; Arthur, J.; Almeida, J. S. *Bioinformatics* **2005**, *21* (9), 1754–7.
- (238) Le Novere, N.; Finney, A.; Hucka, M.; Bhalla, U. S.; Campagne, F.; Collado-Vides, J.; Crampin, E. J.; Halstead, M.; Klipp, E.; Mendes, P.; Nielsen, P.; Sauro, H.; Shapiro, B.; Snoep, J. L.; Spence, H. D.; Wanner, B. L. *Nat. Biotechnol.* **2005**, *23* (12), 1509–15.
- (239) Prieto, C.; De Las Rivas, J. *Nucleic Acids Res.* **2006**, *34*, W298–302 (Web Server issue).
- (240) Bader, G. D.; Hogue, C. W. *Nat. Biotechnol.* **2002**, *20* (10), 991–7.
- (241) von Mering, C.; Krause, R.; Snel, B.; Cornell, M.; Oliver, S. G.; Fields, S.; Bork, P. *Nature* **2002**, *417* (6887), 399–403.
- (242) Rinner, O.; Mueller, L. N.; Hubalek, M.; Muller, M.; Gstaiger, M.; Aebersold, R. *Nat. Biotechnol.* **2007**, *25* (3), 345–52.
- (243) Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* **1999**, *20* (18), 3551–67.
- (244) Geer, L. Y.; Markey, S. P.; Kowalak, J. A.; Wagner, L.; Xu, M.; Maynard, D. M.; Yang, X.; Shi, W.; Bryant, S. H. *J. Proteome Res.* **2004**, *3* (5), 958–64.
- (245) Fridman, T.; Razumovskaya, J.; Verberkmoes, N.; Hurst, G.; Protopopescu, V.; Xu, Y. *J. Bioinf. Comput. Biol.* **2005**, *3* (2), 455–76.
- (246) Sadygov, R. G.; Yates, J. R., III, *Anal. Chem.* **2003**, *75* (15), 3792–8.
- (247) Tabb, D. L.; Fernando, C. G.; Chambers, M. C. *J. Proteome Res.* **2007**, *6* (2), 654–61.
- (248) Higdon, R.; Kolker, E. *Bioinformatics* **2007**, *23* (3), 277–80.
- (249) Titulaer, M. K.; Siccama, I.; Dekker, L. J.; van Rijswijk, A. L.; Heeren, R. M.; Sillevius, S. P. A.; Luider, T. M. *BMC Bioinf.* **2006**, *7*, 403.
- (250) Babnigg, G.; Giometti, C. S. *Proteomics* **2006**, *6* (16), 4514–22.

AC070741J

