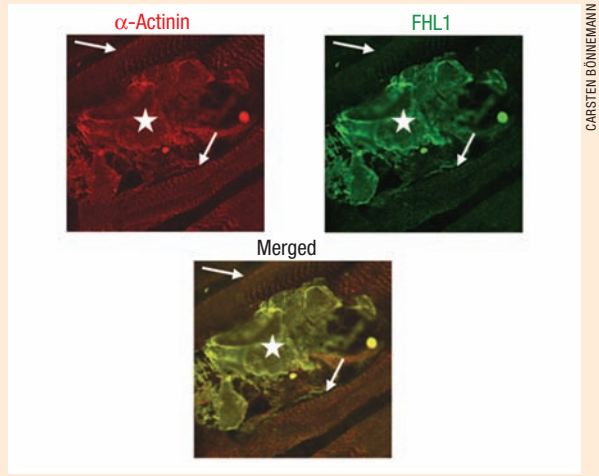
currents Journal of proteome research

Proteomics approach identifies disease gene

Reducing body myopathy (RBM) is a rare muscle disorder characterized by progressive muscle weakness, which often culminates in respiratory failure and death. Although RBM was first described by physicians >30 years ago, its etiology has remained mysterious. Intracytoplasmic inclusion bodies form in the myofibers of RBM patients, but the composition of these protein aggregates and their role in the disease are unclear. Because RBM occurs rarely and sporadically, genetic linkage analysis could not be conducted. That's why Carsten Bönnemann at the Children's Hospital of Philadelphia and co-workers at several universities, hospitals, and institutes in the U.S. and Europe adopted a proteomics approach to identify the RBM disease gene.



FHL1 localization. In muscle fibers with large inclusion bodies (star), FHL1 is depleted from the contractile apparatus (arrows), in which Z disks are labeled with α-actinin.

The inclusion bodies found in RBM patients exhibit many features of aggresomes, or structures that form when the ubiquitin–proteasome system becomes overloaded with unfolded proteins. To investigate which proteins are enriched in RBM inclusions, the researchers used laser microdissection to capture material from inclusion bodies in the muscle biopsies of two female RBM patients. Then, the samples were analyzed by nanoflow LC/MS/MS to identify the most abundant protein components. In addition to known members of RBM aggregates, such as ubiquitin and desmin, a protein called four-and-a-half-LIM-domains protein 1 (FHL1) was enriched in the inclusion bodies of RBM patients.

FHL1, which has an N-terminal zinc finger in addition to 4.5 LIM domains, is encoded by Xq26.3 and is thought to be involved in cytoskeletal scaffolding and transcription-factor regulation. Sequencing of genomic DNA and RT-PCR products from skeletal muscle of the two RBM patients in the study revealed heterozygous mutations in an FHL1 protein region that is essential for zinc binding. Two additional mutations in the same region were found in two other RBM patients.

When Bönnemann and colleagues transfected kidney COS-7 and skeletal muscle C2C12 cells with mutant FHL1 constructs, inclusion bodies formed that contained both mutant and wild-type FHL1, in addition to other muscle proteins and components of the protein-degradation machinery. These results suggest that mutant FHL1 induces the formation of inclusion bodies and that these aggregates sequester wild-type FHL1 and other proteins important for normal muscle function. The researchers say that this study represents the first direct identification of a disease gene by a proteomics approach. (*J. Clin. Invest.* **2008**, *118*, 904–912)

A sweet workflow for glycosylation analysis

Glycosylation, or the covalent attachment of sugar molecules (glycans) to proteins, is an important posttranslational modification (PTM) that affects protein stability, interactions, and activity. But because of the large repertoire of glycans in cells and the heterogeneous glycosylation patterns that can exist for a given glycoprotein, MS analysis of sugar-modified proteins is a sticky situation. To maximize glycopeptide detection in MS, Heather Desaire and co-workers at the University of Kansas optimized each step in the MS analysis of model glycopeptides and applied their strategy to a glycoprotein with 27 glycosylation sites.

First, the researchers compared three sample preparation methods (reversed-phase HPLC [RPLC], Sepharose hydrophilic affinity chromatography, and lectin affinity isolation) for the model glycoproteins immunoglobulin G (IgG), transferrin, and α1-acid glycoprotein (AGP). For IgG and transferrin, which have one and two glycosylation sites, respectively, RP-HPLC fractionation yielded MALDI MS spectra with the highest S/N and the most identified glycopeptides. For AGP, which has five glycosylation sites, a two-step sample preparation with both Sepharose chromatography and RP-HPLC provided the best glycopeptide coverage.

The researchers found that performing MALDI TOFMS in both negative- and positive-ion modes was crucial for the detection of acidic and neutral glycopeptides in the same sample. To help detect weakly ionizing glycopeptides that coeluted with strongly ionizing species, the investigators deglycosylated glycopeptides with peptide-*N*-glycosidase F (known as PNGase F) and confirmed the presence of the coeluting peptides. In the MS data analysis, the number of assigned glycopeptide peaks increased when the team incorporated variable PTMs into the database search. Desaire and coworkers used the optimized method to analyze the heavily glycosylated HIV envelope protein JR-FL gp140 ΔCF. A total of 300 different glycoforms, as well as 23 of the 27 known glycosylation sites of JR-FL, were detected by MALDI TOFMS/MS. The researchers say that these results represent a substantial improvement over standard procedures. (*Anal. Chem.* **2008**, *80*, 3144–3158)

TOOLbox

iCluster

Automated image analysis techniques are often unable to distinguish complex subcellular localization patterns. However, the time required for a cell biologist to examine hundreds or thousands of images in a large data set within such studies is prohibitive. Therefore, Nicholas Hamilton and Rohan Teasdale at the University of Queensland (Australia) developed the iCluster method to visualize and cluster large sets of subcellular localization images. iCluster arranges images in 2D or 3D space so that similar images are close to each other, whereas dissimilar images are far apart. A protein's range of subcellular localizations can be quickly discerned by examination of the clustering pattern. In this way, the cell biologist can judge whether a particular image is truly representative. The researchers tested the algorithm on a collection of 502 images in which 1 of 10 organelles had been stained. iCluster successfully grouped the images into distinct classes of subcellular localization. (*BMC Bioinformatics* **2008**, *9*, 81)

xQuest

Experimental and computational obstacles have hindered analyses of cross-linked proteins by MS. Therefore, to help overcome these difficulties, Ruedi Aebersold and co-workers at the University of Zurich, ETH Zurich, the University of Innsbruck (Austria), and the Institute for Systems Biology devised a method to identify cross-linked peptides from large sequence databases. With the novel strategy, the researchers cross-linked proteins in *E. coli* lysate with an isotopically coded cross-linker. Cross-linked peptides were detected as isotopic pairs in LC/MS spectra and subjected to MS/MS sequencing.

To find candidate peptides, the novel search engine xQuest searches a sequence database with fragment ions common to both partners of an isotopic pair. Then, xQuest enumerates possible peptide–peptide combinations and matches peptide combinations with the precursor ion mass to identify cross-links. The method recognized 22 intraprotein and 7 interprotein cross-links in *E. coli* proteins. xQuest software is available at www.xQuest.org. (*Nat. Methods* **2008**, *5*, 315–318)

Candidate biomarkers in prostate secretions

PSA is a widely used, albeit imperfect, biomarker of prostate disease. The diagnosis and treatment of prostate disease, like many other disorders, would benefit from the availability of a panel of biomarkers. However, the identification of biomarkers in serum is challenging because of the relative scarcity of most of these proteins in blood. In contrast, a bodily fluid near the tissue or organ of interest likely contains a higher concentration of potential biomarkers, which could be identified from the fluid and then screened in serum. To discover candidate biomarkers in the prostate secretions, Biaoyang Lin, Shujie Xia, and co-workers at the Shanghai Institute of Planned Parenthood Research, Shanghai Jiaotong University, Shanghai First People's Hospital, and the Institute for Systems Biology conducted a proteomics analysis of human expressed prostatic secretions (EPS).

EPS sampling, which involves prostate massage and collection of the secreted fluid, is less invasive than a prostate biopsy. Lin, Xia, and colleagues planned to conduct a comparative proteomics analysis of normal and cancerous EPS, but EPS samples proved difficult to obtain from many prostate cancer patients because of disease-related hardening and damage of the gland. Therefore, the researchers decided instead to catalog proteins in EPS samples from a healthy individual, four patients with benign prostatic hyperplasia, and three prostate cancer patients.

Proteomics analysis of human embryonic stem cells

Because human embryonic stem cells (hESCs) can differentiate into many cell types, scientists and the general public have high hopes that hESCs could be used for regenerative medicine or tissue replacement. To realize the potential of hESC applications, researchers must learn more information about the differentiation process. Therefore, Anastasia Yocum and colleagues at the University of Michigan and Applied Biosystems conducted a proteomics analysis of hESCs during induced differentiation.

hESCs. (Left) Phase-contrast micrograph of undifferentiated hESCs. (Right) Fluorescence image of the stained neurofilament of an hESC induced to differentiate into a primitive neuron by noggin treatment. (Image on the left is adapted with permission. Copyright 2008 American Society for Biochemistry and Molecular Biology.)

To keep cells in an undifferentiated state in vitro, hESCs usually require coculturing with irradiated mouse embryonic fibroblasts, but the presence of these feeder cells confounds proteomics analyses. So the researchers cultured hESCs in a defined medium on gelatin-coated plates, with no detectable changes in cell morphology or in the expression of the pluripotency marker OCT3/4. Then, the group conducted a global quantitative proteomics analysis to compare control hESCs with hESCs treated with bone morphogenic protein 4 (BMP4) or its antagonist, noggin. BMP4, an extracellular signaling protein, promoted differentiation toward an epidermal ectoderm lineage, whereas noggin induced primitive neural differentiation. Trypsin-digested proteins from each cell lysate were labeled with isobaric tags for relative and absolute quantitation (iTRAQ) and subjected to 2D LC/TOF/TOFMS.

Compared with untreated hESCs, 34 and 187 proteins were differentially expressed in noggin-treated and BMP4-treated hESCs, respectively. To increase the confidence levels of protein identification and quantification, the researchers implemented a multiple reaction monitoring assay, in which unique (e.g., isoform-specific) peptides and peptide fragments of target proteins were identified in silico and specifically quantified by iTRAQ and LC/TOF/TOFMS. The combined global and targeted proteomics analyses identified several lineage-specific proteins with high confidence. (*Mol. Cell. Proteomics* **2008**, *7*, 750–767)

By 2D microLC/MS/MS, a total of 114 proteins from EPS samples were identified confidently by ≥2 unique peptides. Many of these proteins were shown previously to be prostate-enriched and/or up-regulated at the mRNA level in prostate cancer. Some of the identified proteins (e.g., PSA, KLK2, KLK3) were implicated previously in prostate cancer, whereas others are novel biomarker candidates. Many of the proteins were connected by known protein–protein interactions. According to the authors, this pilot study revealed that EPS is a rich source of potential biomarkers that could aid in the diagnosis and classification of prostate disease. (*Proteomics Clin. Appl.* **2008**, *2*, 543–555)

Improved detection of multiphosphorylated peptides by RPLC/MS/MS

Phosphorylation is an important regulatory mechanism that modulates the activity of many cellular proteins. But the identification of protein phosphorylation sites by MS is notoriously difficult for several reasons. For example, the negative charge conferred by phosphorylated amino acids can decrease a peptide's ionization efficiency during MS in positive-ion mode. The analysis of multiphosphorylated peptides, which have >3 phosphorylation sites, is especially problematic. To improve the detection of multiphosphorylated peptides by RPLC/MS/MS, Zee-Yong Park and colleagues at the Gwangju Institute of Science and Technology (South Korea) reduced the pH of the RPLC elution solvent from ~3.0 to ~1.7. This simple modification decreased the detection limit of a tetraphosphorylated peptide from bovine β-casein by 200-fold.

The researchers observed that under normal pH conditions, tetraphosphorylated peptides had very broad RPLC elution profiles, whereas the same peptides showed typical profiles at low pH. Multiphosphorylated peptides are usually more hydrophilic than their unmodified counterparts. Decreasing the pH of the elution solvent, therefore, could enhance the interaction of multiphosphorylated peptides with the hydrophobic RPLC column material and improve phosphopeptide separation and detection. Under the low-pH condition, more phosphopeptides and proteins were identified in a phosphopeptide mixture from mouse brain tissue. (*Anal. Chem.* **2008**, *80*, 3007–3015)

Do cell phones alter the skin proteome?

For years, scientists and consumers have questioned whether excessive cell phone use could cause physiological changes such as cancer. Although a definitive link to cancer or other diseases has not been proven, Dariusz Leszczynski and colleagues at the Radiation and Nuclear Safety Authority (known as STUK; Finland) demonstrated that the electromagnetic radiation produced by mobile phones alters protein expression in human skin. The researchers irradiated a small area of skin on the forearms of 10 female volunteers with radio-frequency-modulated electromagnetic fields (RF-EMF) for 1 hour. Then, punch biopsies were obtained from irradiated and nonexposed forearm skin of each volunteer. Proteins from the skin samples were separated by 2DE, and the gel was silver stained. Leszczynski and colleagues analyzed the spot distribution pattern in a specific region of each 2D gel (p*I* 4–7, molecular weight <40 kDa). Of the 579 protein spots in this region, eight proteins were differentially expressed between the irradiated and nonexposed skin samples in at least four volunteers. Two of the protein spots were differentially expressed in all 10 volunteers. The researchers say that this finding indicates some reproducibility in the proteome changes. Although larger and more detailed investigations must be conducted, this pilot study suggests that RF-EMF from cell phones can induce changes in human gene expression. (*BMC Genomics* **2008**, *9*, 77)

JUPITERIMAGES

Cellular changes. Can too much yakking on a cell phone alter skin protein expression?

TOOLbox

Warp2D

Accurate comparison of peaks across multiple LC/MS runs is essential for many proteomics and metabolomics experiments, but the LC retention time of an analyte can vary slightly. To correlate peaks among LC/MS runs, most alignment approaches consider only chromatographic retention time. These 1D methods are often unsuitable for complex biological samples with many overlapping peaks. Therefore, Frank Suits and colleagues at the IBM T. J. Watson Research Center and the University of Groningen (The Netherlands) developed Warp2D, a method that uses 2D data (retention time and *m/z*) to align peaks in the time domain. A list of 2D peaks is extracted from each data set, and Warp2D measures the similarity in peaks by an overlap integral. Peaks are aligned in the time domain by correlation-optimized warping. The tool merges the high performance of 1D methods with the robustness of 2D alignment approaches. When applied to several data sets of variable complexity, Warp2D was rapid and accurate. (*Anal. Chem.* **2008**, *80*, 3095–3104)

PeakSelect

In Q-TOFMS/MS spectra of complex protein samples, the distinction of "real" peaks from noise and from isotope peaks of the fragment ion can be problematic. So Jingfen Zhang, Wen Gao, and co-workers at the Chinese Academy of Sciences and the University of Western Ontario (Canada) introduced a preprocessing method, PeakSelect, that helps define real peaks before database searching. PeakSelect identifies a baseline intensity level of noise peaks in the spectra. In addition, the method incorporates an isotope pattern vector that distinguishes noise peaks, which don't have isotopes, and estimates the patterns of isotope peaks for fragment ions. PeakSelect classifies peaks into categories such as noise, single ion peaks, and overlapping peaks. When PeakSelect was used to preprocess spectra from yeast whole-cell lysate, the method not only decreased the Mascot search time but also identified more peptides with improved reliability. (*Rapid Commun. Mass Spectrom.* **2008**, *22*, 1203–1212)