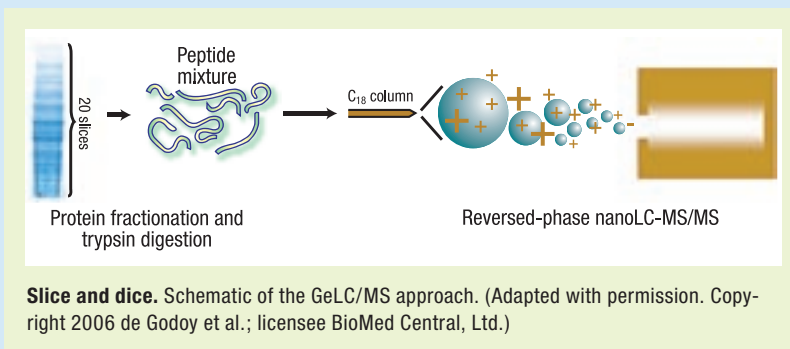


Why can't we analyze a whole proteome by MS?

Try as they might, researchers just haven't been able to elucidate an entire proteome of a cell or an organism with current MS techniques. Smaller subproteomes (e.g., of organelles) that comprise ~1000 proteins have been well characterized, but researchers have found it difficult to go beyond this number for larger samples. So, Matthias Mann and colleagues at the Max Planck Institute of Biochemistry (Germany) and the University of Southern Denmark tried to figure out why. They conclude that the effective dynamic range and the sequencing speed of the mass spectrometer limit the sensitivity of MS for proteomics studies.



In an approach called GeLC/MS, the researchers ran yeast proteins on a 1DE gel and cut the gel into 20 slices. They digested the proteins and analyzed the peptides by reversed-phase nanoLC and MS/MS with a hybrid ion trap-FT mass spectrometer. The researchers analyzed >200,000 spectra and identified 2003 proteins. Although this number was higher than those obtained by other investigators with MS, it still wasn't as many proteins (>4500) as identified in 2 previous studies with non-MS methods (*Nature* **2003**, 425, 686–691; 737–741).

For protein identification, the researchers used stringent cutoff values at which the false-positive rate would be <1 in 10,000. They also tested for false positives by searching for

the proteins in a database that included reversed sequences. None of the identified proteins matched the reversed sequences. Therefore, the researchers say that they have essentially eliminated all false positives.

To understand the limitations of MS as a proteomics technique, Mann and colleagues performed an in-depth examination of their own data. They could detect some, but not all, of the proteins that were present in low copy numbers. Had members

of the rarest class of proteins (in the femtomole range) been present as single proteins, the researchers could have detected them by GeLC/MS. But as members of a complex mixture, most of these proteins were not analyzed. Therefore, the "effective sensitivity" was orders of magni-

tude too low to cover the entire yeast proteome. Likewise, the "effective dynamic range" was orders of magnitude too low.

To determine whether the sequencing speed of the mass spectrometer influenced the effective sensitivity, the researchers conducted a stable-isotope labeling with amino acids in cell culture (known as SILAC) experiment. Often, only one of the pair of labeled peptides was detected by MS; this result suggested that the sequencing speed of the instrument was a limiting factor for the analysis of a complex proteome. The researchers say that the sequencing speed and the effective dynamic range can be improved with better acquisition software and different ion accumulation methods, respectively. (*Genome Biol.* **2006**, 7, R50)

Finding palmitoyl proteins

Many cellular proteins, such as G proteins and some nonreceptor tyrosine kinases, are tethered to the cytoplasmic sides of cellular membranes by a reversible posttranslational modification called palmitoylation or protein S-acylation. Although palmitoylation is a common lipid modification on proteins, little is known about it. For example, consensus sequences that would predict which proteins become modified with palmitoyl residues have not been described in the literature. Current methods for palmitoyl identification are time-consuming and low-throughput.

To date, only 15 proteins of the yeast *Saccharomyces cerevisiae* are known to be modified by palmitoylation. So, Nich-

olas Davis and co-workers at the Wayne State University School of Medicine, the Scripps Research Institute, Michigan State University, and the University of Chicago conducted a proteomics study in *S. cerevisiae* to find new palmitoyl proteins. The researchers identified most of the known palmitoyl proteins as well as 35 new ones.

Davis and co-workers isolated yeast membrane proteins and performed an acyl-biotinyl exchange (ABE) reaction on half of the protein sample; in this reaction, the acyl group is replaced with a biotin moiety. An ABE reaction was performed on the other half of the extract, but a key step was omitted to produce an unlabeled control sample. Streptavidin was used to purify biotinylated proteins.

Proteins in the experimental and control samples were identified by multidimensional protein identification technology (known as MudPIT). The researchers used per-protein spectral counts to determine the relative amounts of each protein.

Of the 15 known palmitoyl proteins, 12 were present at significantly higher levels in the experimental sample than in the control. New palmitoyl proteins, including 8 SNARE proteins that are involved in vesicle fusion, also were identified. By analyzing the proteomes of yeast with gene deletions for palmitoylation enzymes, the researchers determined which proteins were likely substrates of these enzymes. (*Cell* **2006**, 125, 1003–1013)

TOOLbox

PROMPT for comparisons

After identifying proteins by MS, researchers often want to know whether the list includes proteins with certain functions. To perform this analysis, they must integrate sequence and annotation data from many sources that may not be compatible. So, Dmitrij Frishman and Thorsten Schmidt at the Munich University of Technology have developed a software system called protein mapping and comparison tool (PROMPT). The tool uses sequence information to match protein entries with different accession numbers. Protein properties, such as pI or functional categories, are compared, and statistical analyses are conducted. For example, researchers used PROMPT to examine the structural differences between two sets of *E. coli* proteins. Results are visualized as plots or spreadsheets and can be exported in many formats. Academic users can access PROMPT for free at <http://webclu.bio.wzw.tum.de/prompt>. (*BMC Bioinformatics* 2006, 7, 331)

msInspect

Martin McIntosh and co-workers at LabKey Software, the Fred Hutchinson Cancer Research Center, and the University of Washington have developed a bioinformatics software platform for the visualization and quantitation of LC/MS data that is called MS in silico peptide characterization tool (msInspect). With the new platform, data from experiments involving isotope labeling or label-free methods can be analyzed and compared.

The graphical interface allows users to view their LC/MS data as a heat map, with retention time on the *x* axis and *m/z* on the *y* axis. The brightness of the color on the heat map indicates the signal intensity. Users can zoom in and out of the map and view 1D cross-sectional slices of the data. The researchers applied the algorithms to the quantitative analyses of human serum samples that were labeled with the isotope-coded affinity tag (known as ICAT) method or that were unlabeled. In addition, the programs in the msInspect platform (available at <http://proteomics.fhcrc.org>) were used to distinguish two bacterial strains. (*Bioinformatics* 2006, 22, 1902–1909)

Phosphoproteomics of a signal transduction pathway

Her2, a tyrosine kinase, is overexpressed in 20–30% of human breast cancer tumors. To better understand the Her2 signaling pathway, Akhilesh Pandey, Philip Cole, and co-workers at the Johns Hopkins University School of Medicine and the Institute of Bioinformatics (India) have used stable-isotope labeling with amino acids in cell culture (known as SILAC). The researchers compared cells that overexpressed Her2 with control cells and found ~200 proteins with increased phosphorylation and 81 proteins with decreased phosphorylation.

The researchers grew 3 sets of NIH 3T3 cells in different media. Control cells that had been transfected with a vector were grown in regular arginine-containing media, cells that overexpressed Her2 were grown in “heavy” arginine media ($^{13}\text{C}_6$ $^{15}\text{N}_4$ -Arg), and cells that overexpressed Her2 and had been treated with a Her2 inhibitor were grown in “medium” arginine media ($^{13}\text{C}_6$ -Arg). To discover whether Her2 overexpression or inhibitor treatment affected the phosphorylation of proteins, the researchers mixed the cell lysates and isolated the phosphotyrosine-containing proteins with antibodies against phosphotyrosine.

The proteins were run on a 1DE gel and then digested and analyzed by LC/MS/MS.

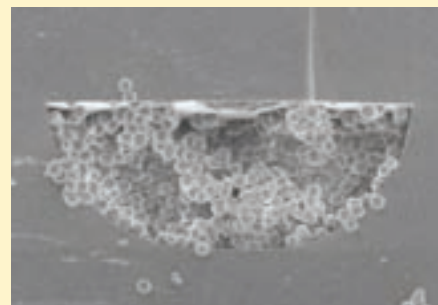
A total of 462 proteins were identified and quantified. Of the 198 proteins that had increased phosphorylation in the overexpressing cells, many were known to be involved in Her2 signaling. In addition, several proteins, such as Stat1, Dok1, and δ -catenin, were identified as Her2 signaling proteins for the first time. The phosphorylation changes of some of these proteins also were verified by western blotting.

The researchers used 2 methods to figure out the roles of the identified proteins in the Her2 pathway. In one approach, they obtained protein–protein interaction data from the Human Protein Reference Database (known as HPRD) and found that 2 of the proteins directly linked the Her2 and epidermal-growth-factor receptor pathways. The researchers also used Bayesian networks to model the Her2 pathway on the basis of the data obtained in this study and 3 additional studies. The modeled network was similar to the known pathway and enabled the researchers to predict how other proteins may fit into the pathway. (*Proc. Natl. Acad. Sci. U.S.A.* 2006, 103, 9773–9778)

A step toward proteomics on a chip

Iulia Lazar and colleagues at the Virginia Polytechnic Institute and State University have developed a microfluidic LC system for proteomics. The new device includes all of the components necessary for separations, including a separation channel, micropump, valve, and ESI interface for MS analyses.

Lazar and colleagues compared the performance of the chip with that of a benchtop micro-HPLC system by running a strong-cation-exchange fraction of an extract from breast cancer cells on both and analyzing the separated peptides by MS. In an initial pass, the researchers identified ~10 \times fewer proteins with the microfluidic system. But when the column length, buffer composition, and sample-injection volumes of the benchtop system were



Small separations. Scanning electron microscopy image of a cross section of a microfluidic LC channel packed with 5- μm particles.

adjusted to match those used with the chip, a similar number of identifications were made. All five potential biomarker proteins that were previously identified with the benchtop system also were found with the LC chip. The researchers plan to develop the microfluidic system into a disposable lab-on-a-chip device. (*Anal. Chem.* 2006, 78, 5513–5524)

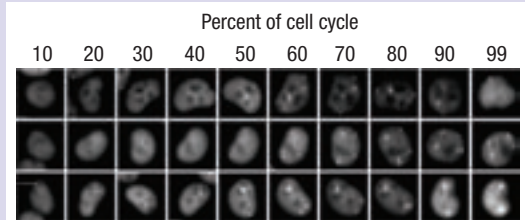
Dynamic proteomics of single cells

The cells in a tissue may look the same at first glance, but interesting protein differences could exist at the single-cell level. Individual cells within a population may not express the same proteins at the same sub-cellular locations or in the same amounts.

To catch some of these differences, Uri Alon and colleagues at the Weizmann Institute of Science (Israel), Stanford University, the Whitehead Institute for Biomedical Research, and Harvard Medical School developed a dynamic proteomics approach. With automated time-lapse microscopy and image-analysis techniques, they observed changes in protein localization in single mammalian cells throughout the cell cycle. In this proof-of-principle study, known and previously unknown changes in protein localization were observed.

The researchers created a library of fluorescently tagged human cells. A retrovirus was used to place the yellow fluorescent protein (YFP) gene into various chromosomal locations in lung carcinoma cells. A total of 200 clones that expressed different YFP fusion proteins at endogenous levels were produced.

Alon and colleagues chose 20 clones in which the YFP-tagged protein localized to the nucleus. They recorded movies of each clone as its cells divided for



Proteins on the move. USP7, a ubiquitin-specific nuclear protease, is expressed in 3 aligned cells. Early in the cell cycle, USP7 is diffusely localized to the nucleus. Later on, the protein moves into nuclear bodies.

RON MILO AND ALEX SIGAL

48–60 h and analyzed the results with automated imaging tools. But instead of synchronizing cells with chemical or mechanical methods that could have physiological effects, the researchers synchronized the cells in silico with a computer program that aligned images of cell-cycle events for various cells. For each cell, they determined the average rates of change of the concentrations of various proteins in the nucleus (nuclear accumulation rates). In addition, the subcellular locations of the YFP-tagged proteins throughout the cell cycle were recorded and analyzed. Novel cell-cycle-dependent changes were observed for some proteins, and the localization of other proteins was found to be variable from one cell to another. The researchers say that this protocol can be applied to the study of proteome changes during any cellular event with an obvious beginning and end. (*Nat. Methods* 2006, 3, 525–531)

Cancer proteins are connected differently

Networking is important when you're on a job hunt. The more connections you have, the better the chance that you'll hear about a great job opportunity. It turns out that increased connectivity also may help cancer proteins to wreak havoc. Paul Bates and Pall Jonsson at the Cancer Research U.K. London Research Institute used a systems approach to determine whether proteins whose genes are mutated in cancer had different connections than noncancer proteins in a human protein–protein interaction network. They observed that cancer proteins interact with a larger number of other proteins and are more often centrally positioned at hubs than noncancer proteins.

A human interactome was constructed on the basis of the interactions of homologous proteins from other species. After scoring each protein–protein interaction, the researchers determined that cancer proteins had 2× more partners than noncancer ones. Another characteristic of cancer proteins was that they appeared in the most tightly connected communities of interacting proteins with a larger number of members. In addition, cancer proteins were more likely to be in overlapping communities. According to the researchers, proteins that play a role in cancer may be older than noncancer proteins in evolutionary terms. (*Bioinformatics* 2006, doi 10.1093/bioinformatics/btl390)

Rapid peptide identifications

Most peptide identification algorithms search experimentally obtained mass spectra against large databases that contain theoretical spectra. Because these searches can take hours, the throughput of proteomics experiments is limited. Databases that include the theoretical spectra of posttranslationally modified proteins can take even longer to search with conventional algorithms. Therefore, Smriti Ramakrishnan and colleagues at the University of Texas, Austin, have developed a filtering method that speeds peptide identifications. The method includes a fast coarse-filtering step in which all true positives and some false positives are obtained. In the second step, a fine filter ranks these data by a significance score. The highest-ranked candidate for each spectrum should be the correct match.

The researchers tested the method with spectra from various protein mixtures and databases containing theoretical *E. coli* and/or human spectra. In every test, the new method correctly identified the proteins in the mixture. According to the researchers, the speed of the new method should allow users to automatically search for mutations and modifications in a high-throughput mode. (*Bioinformatics* 2006, 22, 1524–1531)

MeMo for metabolomics data

Irena Spasić, Douglas Kell, and co-workers at the University of Manchester and the University of Wales (both in the U.K.) have developed MeMo, a model for the representation of metabolomics data and associated metadata. The model includes an administrative module for procedural information, such as protocols and a listing of the instruments that were used. The “wet experiments” components include biological and chemical information about the sample as well as the data generated by analytical instruments. The “dry experiments” module includes the processing and analysis of experimental results. The background knowledge component stores information from external sources in addition to the experimental data, to facilitate the integration of these data sets. MeMo is available at <http://dbkgroup.org/memo>. (*BMC Bioinformatics* 2006, 7, 281)