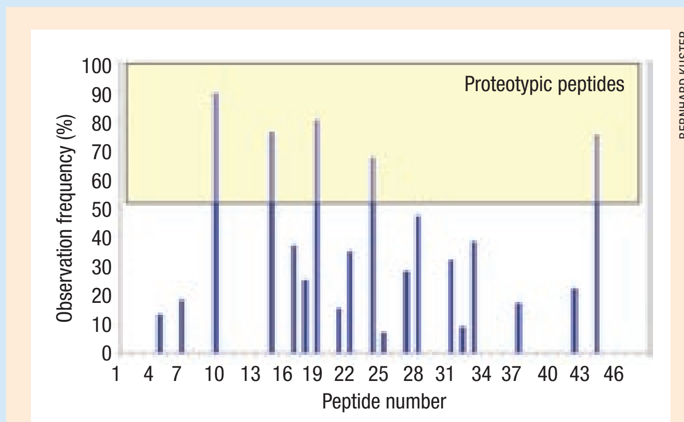
Proteotypic peptides in the spotlight

Some peptides generated by the digestion of a given protein are just detected more often and more reproducibly than others. These proteotypic peptides can provide researchers with valuable information. For example, proteins can be identified with the use of libraries that contain proteotypic peptide sequences. In back-to-back articles, two research groups describe their work on these peptides. Ruedi Aebersold, Bernhard Kuster, and co-workers at the Institute for Systems Biology, Cedars-Sinai Medical Center, the University of California Los Angeles, Cellzome AG (Germany), ETH Zurich, and the University of Zurich developed a method to predict which peptides are proteotypic for a particular protein. In another study, Edward Marcotte and co-workers at the University of Texas used proteotypic peptides to determine absolute quantities of proteins in multidimensional protein identification technology (MudPIT) proteomics experiments.

To predict which peptides are proteotypic for a protein, Aebersold, Kuster, and co-workers studied the physicochemical properties of known proteotypic peptides. Large-scale, well-characterized yeast proteomics data sets that were generated by four common proteomics platforms (1DE/ESI, 1DE/MALDI, MudPIT/ESI, and MudPIT-ICAT) were studied. A total of 494 physicochemical properties, such as charge, likelihood of forming secondary structures, and hydrophobicity, were assessed. The properties that best predicted whether a peptide would be proteotypic varied according to platform. The predictors for yeast also correctly predicted proteotypic peptides for a human data set. The researchers estimate that with their predictors, at least one proteotypic peptide can be identified for most yeast and human proteins. Yeast and human predictors are publicly available and can be downloaded at www.peptideatlas.org. (*Nat. Biotechnol.* **2007**, *25*, 125–131)



You again? A protein's proteotypic peptides are those that are observed more often than the rest.

In the other study, Marcotte and co-workers developed a method for large-scale absolute protein expression (APEX) measurements on MudPIT data. No labels and no added standards are necessary. The algorithm estimates the concentration of a protein in a sample on the basis of the number of its peptides that are detected. Factors such as ionization efficiency and amino acid composition are used by APEX to correct for the fact that some of a protein's peptides are not proteotypic and will not be detected. After the correction is made, the fraction of peptides contributed by a protein to the injected sample pool becomes proportional to the fraction of its peptides that are observed in the MudPIT experiment. The protein abundances calculated with APEX correspond to known abundances over ~2.5 orders of magnitude, are highly reproducible, and correlate well with data obtained with other approaches, such as western blotting. (*Nat. Biotechnol.* **2007**, *25*, 117–124)

Proteomics of the secretory pathway

John Bergeron and co-workers at McGill University (Canada), the University of Montreal, and Göteborg University (Sweden) have developed a quantitative proteomic map of the endoplasmic reticulum (ER) and Golgi apparatus in rat liver. The researchers identified >1400 proteins in the secretory pathway, including 345 that were previously uncharacterized.

Two ER fractions were isolated: rough microsomes (RM), which contain ribosomes, and smooth microsomes (SM), which lack these assemblies. Golgi fractions also were isolated from rat liver cells. Proteins from these organelles were run on 1DE gels. Each lane was sliced into smaller pieces, and the proteins were digested. The peptides were analyzed by MS/MS. To quantify the proteins present in each type of organelle, the researchers applied a redundant-peptide counting approach in which the number of times a peptide is identified in MS/MS spectra is used as an indication of the abundance of the protein from which it is derived.

On the basis of the peptides identified in the study, the researchers concluded that most of the proteins in the RM were involved in protein synthesis and folding. The SM contained proteins in this functional category in addition to detoxification proteins. Most of the Golgi proteins were annotated as soluble biosynthetic cargo. A total of 832 proteins were unique to the ER, 193 were unique to the Golgi, and 405 proteins were found in both organelles.

In addition to the ER and Golgi studies, the researchers prepared coat protomer I (COPI) vesicles, which move from the Golgi to the ER. The role of these vesicles remains controversial. Consistent with a model in which COPI vesicles are involved in the maturation of the Golgi apparatus, the vesicles contained Golgi-resident proteins. (*Cell* **2006**, *127*, 1265–1281)

© 2007 American Chemical Society

TOOLbox

Salvaging true positives from "one-hit wonders"

Recently, scientists have suggested that in a shotgun experiment, only those proteins identified by ≥2 peptides should be reported. However, some proteins identified by one peptide could be correct. The hard part is figuring out which of these "one-hit wonders" are true positives. So, Eugene Kolker and Roger Higdon at the BIATECH Institute and the University of Washington developed a method to distinguish true identifications among single-hit proteins.

In the first step of the method, MS/MS data sets are searched against a database of randomized sequences. Peptides that pass this test are processed with the logistic identification of peptide sequences (known as LIPS) model to obtain the probabilities that peptides are correct on the basis of certain parameters. Cross-validation is the final step. The researchers applied the method to data from three bacterial samples. With the method, they recovered 68–98% of the correct single-hit proteins. (*Bioinformatics* **2007**, doi 10.1093/bioinformatics/btl595)

De novo peptide identification with PILOT

Current de novo sequencing algorithms require large amounts of computing power to generate thousands of candidate sequences. So Christodoulos Floudas and Peter DiMaggio at Princeton University developed an algorithm called peptide identification via integer linear optimization and tandem MS (PILOT) that provides the optimal solution without calculating every possible sequence. A preprocessing step determines which spectral features will be incorporated into the sequencing calculations. Sequences are derived by the use of integer linear optimization. PILOT takes into account single amino acid weights from each peak in a spectrum. If a spectrum is missing some peaks, however, the combined weight of any two amino acids is used instead. In a postprocessing step, candidate peptide sequences are scored. PILOT outperformed five other de novo sequencing algorithms when applied to Q-TOF and ion trap MS data. (*Anal. Chem.* **2007**, *79*, 1433–1446)

Preparing urine samples for metabonomics studies

Urine is a commonly used biofluid in metabonomics studies, but some routine handling steps can be detrimental. For example, manipulations that alter the concentration, pH, or ionic strength of a sample could cause chemical-shift variations in ^{1}H-NMR spectra. So Claus Cornett and co-workers at the Danish University of Pharmaceutical Sciences tested several conditions to determine the best way to prepare urine specimens.

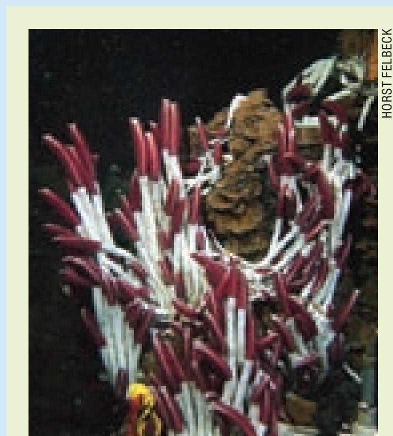
When Cornett and co-workers studied highly concentrated urine samples, they found that a phosphate buffer concentration of 1.0 M was necessary to minimize the variability of the resonance of creatinine. For more dilute urine samples, the best results were achieved when the buffer concentration was 0.3 M. Storage time, temperature, and the presence of preservatives also were assessed. Urine was freeze-dried and stored at –80 °C, or frozen and then placed at 4, –25, or –80 °C, with or without preservatives (sodium fluoride or sodium azide) for 26 weeks. Several artifacts were observed in spectra of samples stored at 4 °C.

Freezing, however, did not have a negative effect on the urine specimens, and preservatives were not necessary when the samples were frozen. In fact, the researchers caution against using sodium fluoride because several artifacts were observed in these samples. Because freeze-drying caused several changes to ^{1}H-NMR spectra, this method of storage is not recommended. (*Anal. Chem.* **2007**, *79*, 1181–1186)

Symbiotic bacteria studied without culturing

Riftia pachyptila, or the giant tube worm, has no mouth or digestive system, but instead relies on its relationship with a single species of sulfide-oxidizing bacteria to derive energy. The characteristics of this endosymbiont are a mystery because it has never been successfully cultured outside its host. Nevertheless, Thomas Schweder and co-workers at the Institute of Marine Biotechnology, the Max Planck Institute for Infection Biology, Ernst Moritz Arndt University (all in Germany), Scripps Institution of Oceanography, Woods Hole Oceanographic Institution, the University of California Santa Cruz, and SymBio Corp. have used a proteomics approach to extract information about the physiology and metabolism of *R. pachyptila*'s endosymbiont.

The researchers first isolated the bacterium from the host tissue, sequenced its genome, and created reference maps of the soluble intracellular and membrane-associated proteins. The sulfide oxidation pathway and reverse (reductive) tricarboxylic acid (TCA) cycle enzymes were expressed at particularly high levels. Similarly, in the cellular extracts of isolated bacteria, the activity of all four reductive TCA enzymes was high; this supported the hypothesis that the reductive TCA cycle is an important component of the endosymbiont's metabolism.



HORST FELBECK

Under the sea. A colony of *R. pachyptila* no doubt enjoying an H_2S snack in a deep-sea hydrothermal vent.

The researchers also found that in times of stress, the bacteria can adjust their method of carbon fixation. Under high-sulfide conditions, the sulfide oxidation pathway spots on 2DE gels show an expected intensification, as do enzymes in the traditional CO_2 fixation pathway, the Calvin cycle. Under low-sulfide conditions, the bacteria increase expression of enzymes for the reverse TCA cycle, a more unusual method of carbon fixation. Oxidative stress caused a corresponding increase in alkyl hydroperoxide reductase; this demonstrated that enzyme's important contribution to the bacteria's oxidative stress resistance. (*Science* **2007**, *315*, 247–250)

iTRAQ labeling of intact proteins

Isobaric tags for relative and absolute quantitation (iTRAQ) typically are added to tryptic peptides in quantitative shotgun proteomics experiments. Labeling intact proteins, however, could allow researchers to use a gel-based workflow and may reduce the complexity of a mixture before MS/MS analysis. So, Bettina Warscheid and colleagues at Ruhr-Universität Bochum (Germany) developed a strategy for tagging proteins with iTRAQ reagents.

To test whether iTRAQ-labeled proteins are compatible with gel-based analyses, myoglobin and bovine serum albumin were labeled and run on a 2DE gel. The researchers observed expected shifts in mobility for both modified proteins relative to unlabeled forms. Next, they tested the labeling efficiency of proteins. A mixture of five proteins was labeled with the iTRAQ-114 reagent, then with the iTRAQ-116 reagent under the same conditions. The proteins were run on a 2DE gel, in-gel digested, and analyzed by MS/MS. Only the 114 tag was observed; this finding indicates that proteins were labeled to completion with the 114 tag.

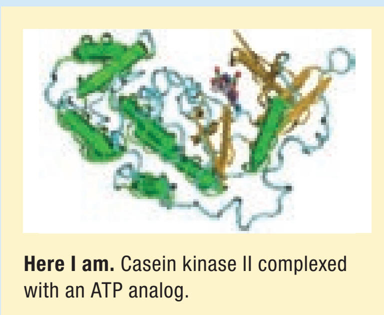
Tryptic peptides only have one site for an iTRAQ tag, but peptides that are derived from a labeled protein may have more than one binding site. To see whether the presence of multiple copies of a label hinders the simultaneous sequencing and quantitation of a peptide, the researchers labeled two aliquots of glucose oxidase with iTRAQ reagents, one aliquot with the 114 tag and the other with the 116 tag, and mixed the samples in various concentration ratios. Typically, labeled peptides are analyzed easily with ESI-MS/MS instruments, such as a Q-TOF. When peptides with multiple tags were analyzed with a Q-TOF, however, sequencing fragments and reporter tags were not observed simultaneously in a mass spectrum. High fragmentation energies were necessary to produce the expected ratios of reporter tags, but at these settings, peptides were broken apart into fragments that were too small for effective sequencing. To solve this problem, the researchers turned to a MALDI TOF-TOF instrument. With this type of mass spectrometer, quantitation and sequence information were obtained simultaneously. (*Proteomics* **2007**, doi 10.1002/pmic.200600422)

Detecting phosphoproteins

The detection and isolation of phosphoproteins remains a challenge for proteomics researchers, so Mary Kay Pflum and Keith Green at Wayne State University have developed a method in which a tag can be easily transferred to a phosphoprotein in a cellular lysate. As a proof of principle, the researchers used the method to label several phosphoproteins with a biotin tag.

Because the ATP used by a kinase typically is partially exposed to solvent at the catalytic site, Pflum and Green theorized that ATP analogs modified at the γ-phosphate could be used to transfer a label to a kinase substrate. To test the hypothesis, they synthesized three peptides with recognition sequences for three kinases. Each peptide was incubated with its kinase and ATP-biotin. All peptides were phosphorylated and biotinylated but with varying efficiencies relative to the modification reactions with unlabeled ATP.

To test the usefulness of the method in biologically relevant samples, cAMP response-element binding protein (CREB) and protein kinase A (PKA) were overexpressed in bacterial lysates. CREB was phosphorylated and biotinylated. In HeLa cells, CREB was modified when ATP-biotin was added, with or without recombinant PKA. Therefore, the reaction can proceed even with endogenous levels of kinase. In addition, recombinant PKA modified several endogenous proteins, so the method could be applied to proteomics experiments as an alternative to radioactive protocols. (*J. Am. Chem. Soc.* **2007**, *129*, 10–11)



Here I am. Casein kinase II complexed with an ATP analog.

TOOLbox

ProCAT for protein microarrays

Although DNA and protein microarrays seem like similar assays, they have different sources of artifacts and noise. Therefore, the algorithms used to evaluate DNA microarrays cannot simply be applied to slides spotted with proteins. So, Michael Snyder and colleagues at Yale University have developed a new protein chip analysis tool (ProCAT) specifically for protein microarrays.

ProCAT has a flexible modular design that includes six components. All components can be run or certain ones can be excluded, depending on the experiment. The background correction module reduces the contribution of smears to the data. In the second module, signals are normalized by a novel method developed for protein microarrays. Another component assigns thresholds for the identification of positive spots. Negative-control and positive-control filters also are included. Protein amounts are normalized with the sixth module. Finally, an annotation report is generated. The entire process takes an average of 5 minutes to complete. (*Genome Biol.* **2006**, *7*, R110)

PRISM for spectral comparisons

Kevin Downard and co-workers at the University of Sydney have developed an algorithm called protein interactions from the spectra of masses (PRISM) as a companion analysis method to a previously reported MS protocol for monitoring complex formation. Although PRISM was originally designed for a specific protocol, the researchers say that it can be applied generally to situations in which the relative intensities of ions in two spectra are compared.

Lists of m/z values and peak areas from MALDI mass spectra are entered into PRISM for the control and reaction samples. A preprocessing step removes the values for those ions that have a low S/N. PRISM's analyzer function compares the areas under the peaks for each ion relative to a third "constant peak" that has the same intensity in both spectra. A graphical user interface enables researchers to visualize the data. (*J. Am. Soc. Mass Spectrom.* **2007**, doi 10.1016/j.jasms.2006.11.005)