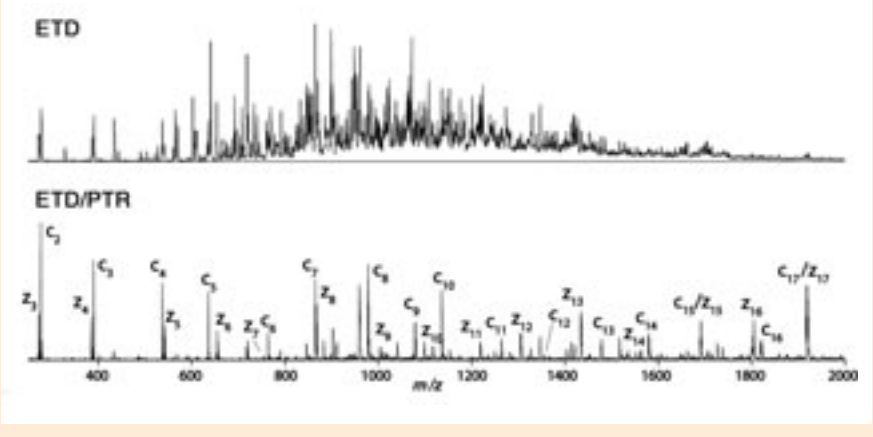
currents Journal of proteome research

Ion/ion reactions for top-down proteomics

With bottom-up MS-based proteomics methods, posttranslational modifications (PTMs) lose their context relative to one another, and alternative splicing variants are often missed. The alternative top-down method, however, requires an expensive Fourier transform MS instrument. To overcome these limitations, Donald Hunt and colleagues at the University of Virginia, Thermo Electron, and the University of Victoria (Canada) have developed a new method that uses both electron transfer dissociation (ETD) and proton transfer reactions (PTR) to obtain N- and C-terminal sequence information from large peptides and whole proteins with a benchtop linear ion trap MS instrument. The addition of PTR to the ETD fragmentation process simplifies the spectra for easier interpretation.

Hunt and colleagues modified a commercially available linear ion trap mass spectrometer with a chemical ionization source that produces anions for ion/ion reactions. Proteins and large peptides are either directly infused into the ion trap or separated by liquid chromatography before MS analysis. Electrosprayed multiply protonated ions react with radical anions of fluoranthene or other polyaromatic hydrocarbons during the ETD process, and the product ions react with benzoic acid anions during a PTR to remove excess charge.



Legible spectrum. PTR simplifies the mass spectrum obtained by ETD for the +13 charge state of ubiquitin. (Adapted with permission. Copyright 2005 National Academy of Sciences, U.S.A.)

The researchers used ETD and PTR to analyze several proteins, including ubiquitin and histones. They sequenced several N- and C-terminal residues for most of the proteins and identified PTMs on the terminal amino acids. The sequences of the middle sections of the proteins were not determined, however, because these fragments lay outside the mass range of the instrument. The researchers believe that optimization of the instrument could resolve this problem. (*Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 9463–9468)

Topology of inner membrane proteins

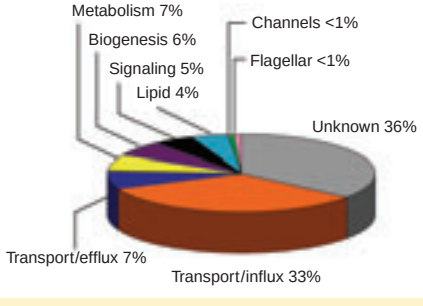
Inner membrane proteins are usually difficult to analyze because of their hydrophobic and amphiphilic properties. To aid future functional analyses of membrane proteins, Gunnar von Heijne and colleagues at Stockholm University and AlbaNova (both in Sweden) investigated the topology of 601 inner membrane proteins in *E. coli*. The results were used to generate topological models of the proteins and provide a global view of the inner membrane.

Topological models describe the orientation and number of transmembrane spans of a protein relative to the lipid membrane. Von Heijne and colleagues previously showed that topological prediction models work better if the C-terminus of a protein is used as a reference point.

To carry out a global topological analysis of the inner membrane proteins in *E. coli*, the investigators used alkaline phosphatase (PhoA) and green fluorescent protein (GFP) as reporter tags for the C-terminus of each of the proteins. PhoA is only active in the periplasm; GFP fluoresces only in the cytoplasm. The combination of the two tags can indicate on which side of the membrane the C-terminus of a given protein is located.

Von Heijne and colleagues cloned 573 genes into an expression vector that attached PhoA and GFP to the proteins' C-termini. An additional 92 genes had one of the tags attached. Out of their initial set of genes, they were able to locate the C-termini of 601 proteins in the bacterial inner membrane.

The investigators sorted the proteins on the basis of their functions and found the largest functional category to be the transport proteins. They point out that their large data set provides a way to overexpress membrane proteins for future functional proteomics and structural genomics studies. (*Science* **2005**, *308*, 1321–1323)



Inner membrane proteome. *E. coli* membrane proteins fall into several functional categories. (Adapted with permission. Copyright 2005 American Association for the Advancement of Science.)

TOOLbox

Proteotypic peptide libraries

To improve the speed of identifying peptide sequences from large sets of MS/MS data, Ronald Beavis and colleagues at Beavis Informatics and the University of Manitoba (both in Canada) have developed a new algorithm called Proteotypic Peptide Profiling (P3). The algorithm relies on a set of proteotypic peptide files, which were derived from a public repository of peptide MS/MS data. The library files were constructed from human and yeast proteomics data collected by multiple laboratories. By using specific sets of peptides rather than every possible peptide that can be generated from a protein sequence, the researchers observed a 20-fold reduction in time needed to assign peptide sequences. The library files, the open-source code for software implementation of P3, and the necessary user interface software are all freely available. (*Rapid Comm. Mass Spectrom.* **2005**, *19*, 1844–1850)

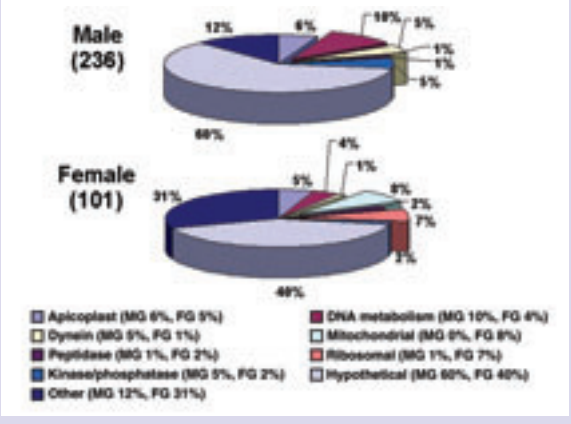
InsPecTing PTMs

Stephen Tanner and co-workers at the University of California, San Diego; the University of Texas Southwestern Medical Center; and the University of Southern California have developed a new tool that identifies posttranslational modifications (PTMs) from MS/MS spectra. Called InsPecT (Interpretation of Spectra with PTMs), the tool relies on a de novo sequencing algorithm to provide sequence tag data. On the basis of the sequence tags, InsPecT filters a database and retains the correct peptide with high probability. An additional InsPecT algorithm that takes peptide fragmentation into account produces a score for identified peptides with PTMs.

The researchers analyzed four data sets with the new tool. InsPecT identified modified and unmodified peptides much faster than Sequest and X!Tandem and assigned many PTMs that were missed by the other algorithms. Researchers can obtain InsPecT by contacting Tanner (stanner@ucsd.edu) or his co-workers. (*Anal. Chem.* **2005**, *77*, 4626–4639)

Proteomes of *Plasmodium* gametocytes

Plasmodium is a one-cell parasite that causes malaria. Andrew Waters, Matthias Mann, and colleagues at the Leiden University Medical Center and the University of Nijmegen (both in The Netherlands) and the University of Southern Denmark recently analyzed the proteomes of the male and female *Plasmodium* gametocytes. The sex-specific proteomes could provide new targets for developing antimalarial vaccines and drugs.



Sex differences. The male (MG) and female (FG) *Plasmodium* gametocytes carry functionally divergent sets of proteins. (Adapted with permission. Copyright 2005 Elsevier, Inc.)

Gametocytes are the precursors to the male and female *Plasmodium* gametes and circulate in a vertebrate host's blood. *Plasmodium* is transmitted from host to host through the bites of female *Anopheles* mosquitoes. The gametocytes develop into the gametes in the mosquito's midgut.

Waters, Mann, and colleagues developed a method to separate and purify the gametocytes. Transgenic *Plasmodium* parasites that expressed green fluorescent protein were sorted by flow cytometry into populations of male and female gametocytes.

Tryptic digests of the proteins from the gametocytes were analyzed by liquid chromatography/MS/MS. The investigators found that the proteomes contained gender-specific proteins. The male gametocyte proteome had 236 out of 650 proteins that were male-specific (36%); the female proteome had 101 out of 541 proteins that were female-specific (19%). Only 69 proteins were common to both sexes.

The proteomes had 29 protein kinases and phosphatases, the majority of which were gender-specific. The investigators studied mitogen-activated protein kinase 2 (MAP2) and NIMA-related kinase (NEK4). MAP2 was found to be necessary for cell differentiation in the male, and NEK4 is needed in the female after fertilization. (*Cell* **2005**, *121*, 675–687)

Changes in saliva biomarkers over time

The abundance of proteins in saliva changes dramatically over the course of a day, according to new research reported by Susan Fisher and colleagues at the University of California, San Francisco, and Applied Biosystems. The results show that a lot still needs to be learned before protein biomarkers in saliva can be used for diagnostic and therapeutic purposes.

The researchers characterized several low-molecular-weight (≤10-kDa) peptides in human parotid saliva, which is known to be rich in peptides. Samples were collected at 3-h intervals during the day, beginning at 9 a.m. and ending at 6 p.m. The samples were collected under conditions that minimized proteolysis, and each sample was derivatized with a different iTRAQ reagent.

Because iTRAQ reagents derivatize primary amine groups, they label all proteins and peptides, except those that lack lysine and reactive N-terminal amino acids. The iTRAQ labeling method allowed the researchers to identify peptides that were not observed in nonderivatized samples. In total, they identified 50 peptides, 19 of which were novel.

The identified peptides were grouped into five categories, depending on the diurnal changes in their concentrations. In some cases, peptide abundance decreased over time; in other cases, it peaked at 12 p.m., 3 p.m., or 6 p.m. In still other cases, peptides showed oscillatory patterns. The results highlight the need to consider diurnal variation as an important variable that influences the composition of saliva. (*Anal. Chem.* **2005**, *77*, 4947–4954)

On-chip protein treatment for SELDI MS

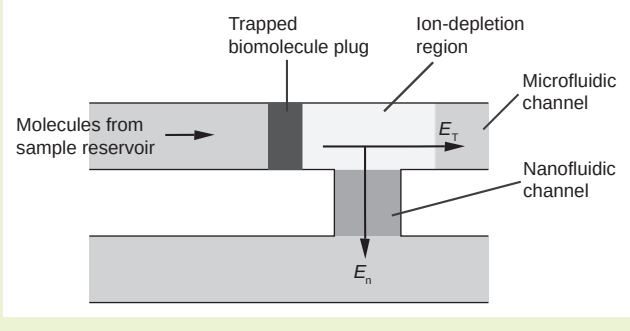
Bernard Gibbs and colleagues at MDS Pharma Services and McGill University (both in Canada) have developed a rapid and simple method for the chemical and enzymatic treatment of proteins. Instead of performing these steps in solution or in a gel, the researchers treat proteins directly on a SELDI chip. The method can analyze low-picomolar amounts of protein, and it can be applied to both pure proteins and mixtures.

To demonstrate the method, the researchers performed separate on-chip deglycosylation and dephosphorylation reactions on three purified proteins. Mass spectra from intact forms of the proteins were compared with those obtained from treated proteins. Both types of reactions yielded forms with masses that were consistent with the loss of oligosaccharides or phosphoric acid groups. The physicochemical properties of the chip's surface affected the reaction rates.

The three proteins were also chemically treated and proteolyzed on-chip in separate reactions. The researchers successfully used data from the peptides to identify the original proteins. In addition, they identified two proteins from an on-chip proteolyzed rat urine sample. (*Anal. Chem.* **2005**, *77*, 3644–3650)

Protein preconcentrator

Although PCR is widely used in genomics to amplify DNA signals, an analogous tool does not exist for proteins and peptides. Proteomics researchers typically rely on preconcentration techniques, which concentrate proteins and peptides from a dilute solution, to boost detection sensitivity. In general, these techniques enhance the signal as much as 5000×.



Concentrate it. A schematic of a nanofluidic protein concentration device. E_n is the electric field applied across the nanochannel, which generates the ion-depletion region, and E_T is the tangential electric field applied across the ion-depletion region.

To strengthen the signals of proteins and peptides even more, Jongyoon "Jay" Han and colleagues at the Massachusetts Institute of Technology developed a simple microfluidic preconcentration device based on electrokinetic trapping by a nanofluidic filter. The new device offers preconcentration factors as high as 10^8.

The new device consists of two microfluidic channels bridged by a buffer-filled 40-nm nanochannel, which acts like an ion-selective membrane. When an electric field is applied across the nanochannel, more counter-ions (positively charged) migrate across it than co-ions. As the applied voltage increases, ion transport becomes diffusion-limited and an ion-depletion zone is generated. Above a certain threshold voltage, electroneutrality is no longer maintained, and the nanochannel develops an extended space charge layer. If a tangential electric field is then applied along the microchannel, a nonlinear electrokinetic flow is induced, and charged biomolecules accumulate in front of the extended space charge layer.

One of the notable features of the device is its speed. The researchers obtained a 10^7-fold increase in concentration in 1 h. Several factors, however, affected its efficiency. Among them were ionic strength of the buffer, pH, the size of the channels, and the number of nanochannels.

The researchers have already coupled the device with on-chip free-solution electrophoresis to demonstrate that it can be turned on and off. The next step is to couple it with MS. (*Anal. Chem.* **2005**, *77*, 4293–4299)

TOOLbox

S-score

Although the Mascot score (M-score) is commonly used to help researchers identify peptides and proteins from database searches of MS/MS spectra, a low M-score does not always indicate a poor match. To overcome this limitation, Mikhail Savitski and colleagues at Uppsala University (Sweden) have developed a new database-independent scoring method called S-score, which relies on the longest peptide sequence tag provided by data from both collisionally activated dissociation and electron capture dissociation.

The new method filters out poor MS/MS data before the database search; this leads to faster data analysis and fewer false positives. In addition, it validates spectra with below-threshold M-scores and high S-scores, and it considers those with high S-scores and zero M-scores (no match) to have modified sequences. According to the researchers, a combination of M-scores and S-scores can positively identify >25% of all MS/MS data. (*Mol. Cell. Proteomics* **2005**, *4*, 835–845)

PROTICdb

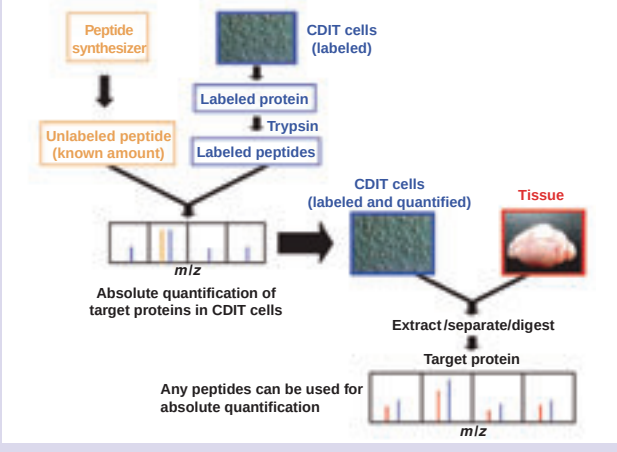
Johann Joets and co-workers at the Centre National de la Recherche Scientifique, the Institut National de la Recherche Agronomique, and the Université Victor Segalen Bordeaux 2 (all in France) have developed a web-based database tool called PROTICdb that stores plant proteome data and records details about methods and conclusions. Although the database does not include analysis tools, data can be exported to third-party statistical software packages. Data are stored as projects, which can be described by a controlled vocabulary of keywords. Access to the information in the database is restricted to database administrators, users, and guests who are specified by users. Although PROTICdb was developed for plant data, it can be adapted for the storage of proteomics information from other organisms. The tool and the source code are freely available at http://moulon.inra.fr/~bioinfo/PROTICdb. (*Proteomics* **2005**, *5*, 2069–2081)

Quantitative proteomics with CDITs

Yoshiya Oda and colleagues at Eisai Co. (Japan) have developed a new approach for quantitative tissue proteomics that relies on culture-derived isotope tags (CDITs) as internal standards. According to the researchers, the CDIT strategy is simple, convenient, and low-cost, and it allows quantification of more proteins than the isotope-coded affinity tag method.

To demonstrate the new approach, the researchers cultured mouse Neuro2A cells in a ^{13}C-enriched medium and mixed them with mouse whole-brain tissue samples. They identified and quantified 1000 proteins, of which at least 97% were expressed in both the mouse whole-brain and Neuro2A cells.

In addition to providing relative concentrations between two samples, the CDIT method can provide absolute concentrations of proteins. The researchers used synthetic unlabeled peptides to quantify the corresponding ^{13}C-labeled proteins in cultured Neuro2A cells. The results were then used to calculate the absolute amounts of 103 proteins in mouse whole-brain samples. The expression levels in the whole-brain samples were in good agreement with those in the Neuro2A cells. (*Nature Biotechnol.* **2005**, *23*, 617–621)



Absolute quantification. The absolute concentrations of proteins can be determined using CDITs. An unlabeled synthetic peptide is used as an internal standard for a target protein that is expressed in CDIT cells. (Adapted with permission. Copyright 2005 Nature Publishing Group.)

Glycoprotein analysis

Jasna Peter-Katalinić and colleagues at the University of Münster and the University Hospital of Münster (both in Germany) have developed an MS method to determine the structures of N-linked glycans attached to proteins as posttranslational modifications. The technique allows for the analysis of glycoproteins by both positive- and negative-ion MS from the same MALDI spot.

In the new method, glycoproteins are excised from 1-D or 2-D gels and deglycosylated by treatment with PNGase F. The researchers used graphitized carbon columns to purify the samples. Most of the glycans were analyzed directly by MS, but sialylated glycans were methylated before MS analysis to avoid spontaneous fragmentation. The researchers achieved the best MALDI results for both positive- and negative-MS modes when they applied the matrix 2,4,6-trihydroxyacetophenone monohydrate mixed with diammonium citrate.

Peter-Katalinić and colleagues used the new method to characterize individual glycans, such as transferrin, in the plasma of patients with congenital disorders of glycosylation (CDG). In addition, they used high-pH anion-exchange chromatography with pulsed amperometric detection before MS to compare the quantities of several glycosylated plasma proteins of CDG patients with those of healthy subjects. (*Proteomics* **2005**, *5*, 2689–2701)

SNPs, alternative splicing, and PTMs

By creating a proteomics-scale database of human proteins that integrates DNA, RNA, and protein-level information, Neil Kelleher and colleagues at the University of Illinois, Urbana–Champaign, have characterized single nucleotide polymorphisms (SNPs), alternative splicing, and posttranslational modifications (PTMs) on intact proteins by Fourier transform MS/MS. The results pave the way for routine use of top-down MS/MS to reveal variants on a large number of human proteins.

One key advantage of the top-down approach is that the entire primary structure of a mature protein can be analyzed. This allows for the concurrent analysis of genotyping SNPs, splice variants, and PTMs—a feat that is not possible with digestion-based approaches.

The researchers used both collisionally activated dissociation and electron capture dissociation to analyze 2–3 small proteins from human HeLa cells. They identified 45 protein forms, which revealed 34 coding SNPs, 2 alternative splicing variants, and 12 diverse PTMs, including a previously unidentified phosphorylation site. In addition, 45% of the proteins were in forms not present in UniProt's Human Proteomics Initiative, and 85% required no manual validation of the search results.

According to the researchers, the new informatics strategy, called shotgun annotation, improves protein identifications, enables the simultaneous characterization of known biological events, and reveals unknown biology more efficiently than other approaches. The overall process, they say, can be thought of as proteotyping, a term that is similar to genotyping at the DNA level but reveals protein variability among populations and over time. (*Mol. Cell. Proteomics* **2005**, *4*, 1002–1008)

The missing selenoproteome

Vadim Gladyshev and co-workers at the University of Nebraska have identified 209 known selenoproteins and 101 previously uncharacterized selenoproteins in microbial sequences obtained by the Sargasso Sea environmental genome project. The results are key to understanding selenium biology.

Selenoproteins are difficult to identify because the DNA codon that signals the incorporation of selenocysteine, a rare amino acid, also signals the end of translation. Most annotation tools miss selenoproteins when the tools are used to predict theoretical proteins from genomic information. Because cysteine often replaces selenocysteine in homologs of most known selenoproteins, the researchers applied an algorithm that compared the microbial sequences with those of homologous genes in other organisms. They also filtered the data on the basis of conserved flanking sequences to improve the algorithm's performance. None of the identified selenoproteins had been annotated correctly by the Sargasso Sea project. (*Genome Biol.* **2005**, *6*, R37)