

Probing the proteome

The completion of the human genome sequence, coupled with analytical techniques such as mass spectrometry, has fuelled interest in proteomics. Diane Gershon reports.

Even before the draft sequence of the human genome was completed, investors and companies were looking to the 'next big thing' — proteomics. Huge amounts of time and money were invested in proteomics-related ventures, and a crop of start-up companies emerged. Some firms set up large-scale proteomics factories running around the clock, with the aim of taking an inventory of the entire human proteome. A few years down the line, reality is setting in. How to turn proteomics data into products remains an open question and investors are becoming impatient.

"I think what it shows is that this is not a simple technology space and the barriers to entry are very high," says Keith Williams, chief executive of Proteome Systems in North Ryde, Australia. But nothing has changed in terms of what people need to do, he says. "Proteins are absolutely at the centre of where the products are going to be." Most drugs currently on the market, or in development, are directed against protein targets.

And the field still offers a major market opportunity for instrument and reagent companies. According to international analysts Frost & Sullivan, the worldwide market for proteomics technologies is projected to



Frank Laukien (right) and Michael Easterling with Bruker Daltonics' \$2-million, 12-tesla FT-mass spectrometer.

more than double between 2002 and 2005, increasing from an estimated US\$2.4 billion to \$5.8 billion.

Proteomics encompasses systematic studies of the identity, changing abundance, distribution, modifications, interactions, structure and function of large sets of proteins, as well as their involvement in disease.

But the technical challenge cannot be

overstated. Multiple proteins can be obtained from each gene, giving vastly more proteins in the human proteome than the 30,000–40,000 genes estimated for the genome. Proteins are also inherently unstable outside a narrow range of environmental conditions. And the wide and dynamic range of protein abundances makes it hard to detect low-abundance proteins against a high-abundance background.

The sheer number of different tasks and objectives in proteomics means that there is unlikely to be a universal technology "that enables you to identify all proteins, in all samples, all of the time", says Kevin Auton, chief executive of proteomics technology developers NextGen Sciences of Huntingdon, UK.

Although equipment tailor-made for proteomics is beginning to appear on the market, access to the technology, some of which comes with a high price tag, is still an issue for many researchers. More automated and integrated proteomics platforms that are both robust and user-friendly are needed, particularly for proteome-wide analyses.

Mass spectrometry (MS) has been key to the development of proteomics. It can be used to identify and, increasingly, quantify large numbers of proteins from complex samples

BRUKER DALTONICS

MASS SPECTROMETRY GOES MAINSTREAM

At the less expensive end of the market is the ion-trap mass spectrometer, which is reasonably sensitive and has a reputation for being rugged, reliable and easy to use. But conventional three-dimensional ion traps have limited trap capacity, although newer instruments offer improvements on this front. The 1100 Series LC-MSD Trap XCT non-linear ion trap from Agilent in Palo Alto, California, for example, has a trap capacity four to eight times greater than its predecessor, which the company says will give a tenfold increase in sensitivity. Prices start at about US\$200,000. The linear or two-dimensional ion-trap machines, such as the LTQ LC-MS/MSⁿ from Thermo Finnigan in San Jose, California, are also configured to offer greater ion storage capacity.

The new 4000 Q TRAP LC-MS/MS system from Applied Biosystems in Foster City, California, and MDS Sciex in Concord, Canada, is a triple-quadrupole mass spectrometer where the last quadrupole can be operated as a conventional quadrupole mass filter or as a linear ion trap — an arrangement that allows decoupling of precursor ion isolation and fragmentation from the ion trap itself. The company says that this provides more sensitive tandem mass-spectrometry (MS/MS) and allows the identification and measurement of low levels of post-translational modification in a single liquid chromatography–MS/MS run. "This is one of the few instruments where you'd actually be able to do the protein identification work, protein characterization and absolute quantitation on the same instrument,"

claims Ron Bonner, director of applied research at MDS Sciex.

At the top end of the market, Bruker Daltonics in Billerica, Massachusetts, and Thermo Finnigan have hybrid Fourier transform MS (FTMS) instruments. Bruker's APEX-Q analyses ions in the ion cyclotron resonance cell under high vacuum in a high magnetic field. Frank Laukien, president and chief executive of Bruker Daltonics, says that the hybrid 'front-end' for FTMS, high-field magnets (9.4 or 12 tesla) and COMPASS software are suitable for 'shotgun' proteomics as well as a 'top-down' approach that does not depend on enzymatic digestion as a precondition for protein-sequence analysis. The Finnigan LTQ-FT consists of a linear ion trap, a high-transmission ion-guidance system, and an ion cyclotron resonance analyser based on a 7-T superconducting magnet.

But the ultra-high resolution and greater mass accuracy of FTMS comes at a price. The APEX-Q (9 T model) starts at about \$1.25 million and rises to \$2 million for the top-of-the-range 12-T model. And FTMS instruments have a reputation for being tricky to operate, although Laukien says that "anybody who can run an ion trap or a Q-TOF can also run a FTMS". Whether the enhanced analytical capability of these high-end instruments will offset their cost remains to be seen. John Yates of the Scripps Research Institute in La Jolla, California, believes it might. "I think these are going to have a huge impact if they wind up being user-friendly," he says.

D.G.

(see R. Aebersold and M. Mann *Nature* **422**, 198–207; 2003). “If you want to just do protein identification in fairly normal organisms, such as yeast, humans and mice, where the genomes are known, then that’s pretty straightforward,” says Ron Bonner, director of applied research at mass spectrometer developers MDS Sciex in Concord, Canada.

Caught in a trap

Mass spectrometers consist of an ion source, a mass analyser and a detector. Of the four main types of mass analyser currently in use (ion trap, time-of-flight, quadrupole, and Fourier transform ion cyclotron), the workhorse has been the standard three-dimensional (3D) ion trap. “They’re relatively cheap, they’re extremely robust and they have good performance, but the mass accuracy and mass resolution is really not all that great,” says Ruedi Aebersold of the Institute for Systems Biology in Seattle, Washington. “Sensitivity-wise, they’re actually very good.”

The more recent linear, or two-dimensional, ion traps (see ‘Mass spectrometry goes mainstream’, page 581) have better trapping efficiency and increased ion capacity, which gives them greater resolution, sensitivity and mass accuracy. Electrospray ionization, which ionizes the analytes out of a solution and so can be coupled to liquid-based separation systems, is used in combination with a variety of mass analysers, particularly ion traps. Time-of-flight (TOF) analysers measure the mass of intact peptides with high accuracy and resolution and are often used for high-throughput protein identification



Keith Williams says that providing start-to-finish solutions is key.

by peptide mapping (peptide mass fingerprinting). In these machines, the ionization method is typically matrix-assisted laser desorption/ionization (MALDI).

The novel Sensitizer peptide-labelling reagent developed by Proteome Sciences of Cobham, UK, is aimed at improving the sensitivity of MALDI-TOF and SELDI (surface-enhanced laser desorption/ionization)-TOF MS systems, in particular, without requiring modifications to instruments or data-analysis software. Increasing the number of proteins identified by peptide mass fingerprinting should mean that fewer samples require full *de*

novo sequencing using the more time-consuming tandem MS (MS/MS).

The buzz this year is over Fourier transform MS (FTMS), with two hybrid machines specifically targeted at proteomics coming onto the market (see ‘Mass spectrometry goes mainstream’, page 581). FTMS was used to identify more than 60% of the predicted proteome of the bacterium *Deinococcus radiodurans*, the most complete coverage of any proteome yet. And although the analysis of post-translational modifications such as glycosylation and phosphorylation using MS is still something of a challenge, FTMS was able to identify almost every post-translational modification of a 29 kDa protein.

The well-established technique of two-dimensional gel electrophoresis (2DE) has been used most commonly for separating protein mixtures in preparation for MS. And this is likely to continue in the near future. To some extent, “it’s because old habits die hard”, says Aebersold. After staining and image analysis, selected spots from the gel are excised, digested and the peptides analysed by MS. Despite attempts at automation (see ‘Look, no hands’, below), 2DE is likely to remain fairly low-throughput. It also requires relatively large amounts of sample (an issue with clinical material) and in the past has failed to detect low-abundance proteins reliably. Pre-fractionation of the sample, or the use of more sensitive staining methods and larger-format gels, can enhance detection of low-abundance proteins.

The shortcomings of 2DE-MS have fuelled interest in eliminating the need for gels. Liquid

LOOK, NO HANDS

A recurring theme in proteomics is the demand for greater automation to increase efficiency and reproducibility. But to improve somebody’s process, “you first have to convince them they’ve got one”, says John Hobbs, strategic market manager for Beckman Coulter in Fullerton, California. Beckman Coulter has modified and repositioned some of its products under the name ProteomeLab. The products are very “solutions-orientated”, says Hobbs. With the PF 2D automated, two-dimensional protein-fractionation system, for example, the company supplies the chemistry, the hardware, the software and the methods, and “we basically tell the customer don’t mess with it”, he says. The PF 2D uses chromatofocusing followed by non-porous reverse-phase chromatography for high-resolution separation of proteins from complex mixtures. Fraction collection from the first dimension is controlled by an in-line pH monitor, with automatic injection of the fractions into the second dimension. Liquid fractions can be stored or the eluent connected directly to electrospray ionization-mass spectrometry. Beckman Coulter is working with mass spectrometer vendors to ensure compatibility with MALDI (matrix-assisted laser desorption/ionization) plate spotters.

Although non-gel-based methods coupled with mass spectrometry (MS) are gaining acceptance, the most commonly used approach is still a combination of two-dimensional gel electrophoresis (2DE) and MS. Two-dimensional gel electrophoresis is difficult to automate, but NextGen Sciences of Huntingdon, UK, will soon be shipping a fully automated 2DE

platform — a2DE. Immobilized pH gradient strips (up to 30 cm) are run in parallel in the first dimension and three large-format SDS-PAGE gels in the second, and all steps between sample injection and SDS-PAGE are automated. Future versions will allow the same degree of optimization in the first dimension as currently offered in the second dimension, says chief executive Kevin Auton. Then “you can really start to zoom in on certain isoelectric points, as well as molecular-weight ranges”, he says.

For about US\$2 million, Proteome Systems in North Ryde, Australia, sells a start-to-finish automated proteomics solution called ProteomIQ, launched last year — or you can buy the components separately. It comprises kits and instruments for sample preparation and protein separation, Shimadzu-Biotech’s Axima CFR Plus MALDI-TOF (time-of-flight) mass spectrometer and bioinformatics capabilities. The Xcise gel-processing robot, developed jointly with Shimadzu-Biotech of Kyoto, Japan, bundles the functions of gel imaging, precision spot cutting, protein digestion and peptide purification, and deposition onto a target plate into one instrument. Also developed with Shimadzu-Biotech, the novel Chemical Inkjet Printer (ChIP) uses piezoelectric printing technology to dispense minute amounts of reagents onto selected protein spots electroblotted from 2D gels onto membranes. After processing, the membrane can be placed directly into the mass spectrometer for peptide analysis. ChIP not only cuts down on liquid-handling steps but also offers the option of archiving samples for future analysis.

D.G.

chromatography (LC)–MS/MS is gaining in acceptance and is generally more amenable to automation. If quantification is required, the proteins or peptides can be labelled with stable isotopes to enable measurement of differential levels of protein expression. One option is the site-specific, covalent labelling of proteins with isotope-coded affinity tags such as the ICAT reagents from Applied Biosystems in Foster City, California.

Technologies for analysing biomolecular interactions, such as surface plasmon resonance (SPR), are also hitching themselves up to MS. BIAcore of Uppsala, Sweden, has integrated its BIAcore 3000 SPR-based instrument with MALDI–MS analysis so that proteins bound to the ligand on the sensor surface can be automatically recovered and deposited directly onto the MALDI plate for further analysis by MALDI-TOF or TOF/TOF–MS.

Good preparation

But no matter what the analytical method, good sample preparation is essential. Fractionation methods such as centrifugal ultrafiltration, immunoaffinity purification and free-flow electrophoresis (FFE) can be used to reduce the complexity of the sample and enrich rare proteins before processing by 2DE or LC–MS. Tecan in Männedorf, Switzerland, sells the ProTeam FFE workstation, a matrix-free FFE fractionation system for processing proteins, organelles, membrane fragments or whole cells. The sample flows continuously in a thin film of aqueous medium between two parallel plates. The application of a high

voltage generates an electric field perpendicular to the laminar flow. Charged species are deflected, allowing fractions to be collected on a fast, preparative and continuous basis.

The Gyrolab MALDI SP1 from Gyros in Uppsala, Sweden, is a microlaboratory in the form of a compact disc on which sample preparation procedures (sample clean-up and concentration) for peptide mapping or sequencing by MALDI–MS are miniaturized and integrated. Gyros also sells the MALDI IMAC for concentrating, purifying and crystallizing phosphorylated peptides directly onto MALDI target areas on the CD. Gyrolab workstations and kits are available for use with MALDI mass spectrometers from Shimadzu-Biotech in Kyoto, Japan, and Bruker Daltonics in Billerica, Massachusetts. MassPREP target plates from Waters-Micro-mass of Milford, Massachusetts, are designed for use with the company's robotic protein-handling system (MassPREP Station) and the Waters MALDI-TOF family of mass spectrometers. They provide fast, simple 'on-plate' preparation of protein digests and are designed to offer a tenfold increase in sensitivity over conventional target plates.

Several companies now sell automated equipment for spotting proteins onto MALDI targets. The Microlab Star MALDI Spotting Workstation from Hamilton Bonaduz, Switzerland, for example, has a volume range of 0.5–1,000 microlitres and can spot a 384-well target in 1 or 2 hours, depending on the number of channels fitted. The Ettan Spot Handling Workstation 2.1

from Amersham Biosciences in Piscataway, New Jersey, automates the process from spot picking from 2D gels through to digesting, drying and spotting the resultant peptides onto the target plates. A sample loader that will automatically place the target plates into the mass spectrometer is in the works.

With many of the component techniques now automated, the next trick is to bring them together. Start-to-finish automated proteomics systems are beginning to appear. Last year, Proteome Systems launched ProteomIQ (see 'When the chips are down', below). Others hope to do likewise. Amersham Biosciences has joined forces with Thermo Electron in Waltham, Massachusetts, and Waters-Micro-mass has partnered with Bio-Rad of Hercules, California, to develop ProteomeWorks.

Alternative angles

A quite different approach to probing protein activity and function is the protein microarray, of which there are two basic types. The analytical microarray contains an ordered array of protein-specific ligands, typically antibodies, spotted onto a derivatized solid surface. They can be used to monitor differential protein expression, protein profiling and clinical diagnostics. But progress here is constrained by a lack of comprehensive sets of high-specificity, high-affinity antibodies.

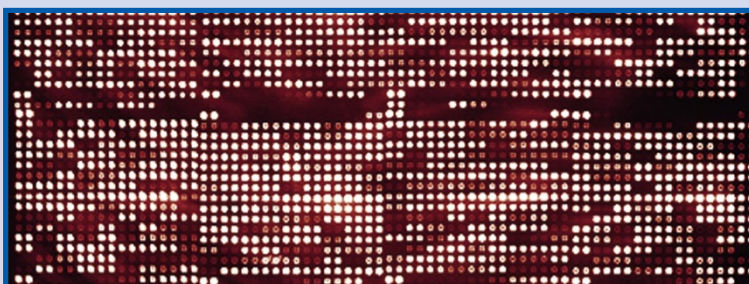
UK-based Cambridge Antibody Technology is tackling this problem by using phage display to create large libraries of antibodies

WHEN THE CHIPS ARE DOWN

Off-the-shelf protein arrays are just beginning to trickle onto the market. But most have been low-density antibody arrays for profiling cytokines, secreted proteins that act as intercellular chemical messengers. Panomics of Redwood City, California, produces TranSignal Human Cytokine Antibody Array 2.0 for profiling the expression of 21 well-studied human cytokines. The array uses standard chemiluminescence detection and detects soluble cytokines at concentrations in the picogram per millilitre range. Zyomyx of Hayward, California, offers a human cytokine biochip for profiling 30 cytokines, which is sold for use with its Protein Profiling Biochip System, which includes an automated workstation, biochip reader and data-analysis software.

BD Biosciences Clontech in Palo Alto, California, sells an off-the-shelf antibody array for detecting a variety of cytosolic and membrane-bound proteins relevant to signal transduction, cell-cycle regulation, gene transcription and apoptosis. Its Antibody Microarray 500 contains over 500 antibodies printed on a standard-size (75 × 25 × 1 mm) glass slide and uses fluorescence-based detection. It is designed for qualitative rather than quantitative analysis but can detect differences in protein abundance between two samples with detection limits of 20 pg ml⁻¹ for each protein target.

On the protein-function microarray front, Protometrix of Branford, Connecticut, will later this year roll out its first product, the Yeast ProtoArray, which stems from the pioneering protein microarray work of Michael



The Yeast ProtoArray from Protometrix contains some 5,000 proteins.

©PROTOMETRIX INC., USED WITH PERMISSION

Snyder's group at Yale University. This 'proteome' microarray will contain some 5,000 proteins in the *Saccharomyces cerevisiae* (yeast) proteome, including over 1,500 with significant homology to human proteins.

"The biggest bottleneck for everybody is still content," says Barry Schweitzer, Protometrix's senior director for technology. The company has now industrialized the cloning, expression and protein-purification process. "It's just a factory running 24/7," he says. "Everything that we clone we also have to sequence." The Yeast ProtoArray is targeted more towards the research market but Protometrix expects to follow this up with products of direct relevance to drug discovery.

D.G.

with the desired specifications. The company is also developing a cell-free system for producing antibodies using 'ribosome display' technology, which may make it possible to build larger libraries than with phage display.

Other companies are exploring capture agents that can be generated more efficiently than antibodies and are more stable to heat or changes in pH. These include protein scaffolds, such as Trinetin binding proteins from Phyllos in Lexington, Massachusetts, and Affibody affinity ligands from Affibody in Stockholm; RNA or DNA aptamers — oligonucleotide sequences — from SomaLogic of Boulder, Colorado; ribozymes (Ribo-Reporters) from Archemix in Cambridge, Massachusetts; and partial-molecule polymeric imprints (ProteinPrint films) from Aspira Biosystems in South San Francisco, California. Commercial protein microarrays are starting to appear on the market, but so far most are relatively low-density antibody arrays for profiling cytokines.

Functional fit

A second type of protein microarray is the functional protein microarray, with potential for the high-throughput analysis of whole proteomes or other large collections of proteins or protein domains. Gavin MacBeath, assistant professor in the department of chemistry and chemical biology at Harvard University, is using such microarrays to help understand how the cell regulates complex processes such as apoptosis, growth-factor signalling, intercellular communication and protein trafficking.

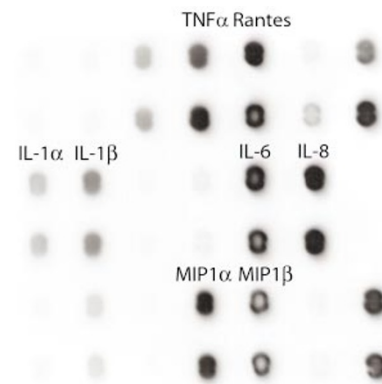
But it is slow going, he says. Because there are no large, cheap, defined sets of cloned genes available, he has to clone the genes and then express and purify the proteins required for the arrays. MacBeath arrays collections of proteins or protein domains onto a variety of chemically derivatized glass slides such as those from TeleChem International in Sunnyvale, California, typically using a home-built arrayer or a machine from Affymetrix in Santa Clara, California. Standard fluorescent dyes are used as labels and slides are scanned using either an ArrayWoRx scanner from Applied Precision in Issaquah, Washington, or a GenePix 4000B microarray scanner from Axon Instruments in Foster City, California.

Most papers to date on protein arrays have been 'proof-of-concept', he says. "We're at the point now where people need to start applying this technology to real questions and use it to learn something about biology."

There are now tools on the market to make array technology more accessible and higher throughput. PerkinElmer Life Sciences in Boston, Massachusetts, for example, sells the ProteinArray Workstation, which automates the processing of multiple protein microarrays. Developed in conjunction with NextGen Sciences, the system can process up



Gavin MacBeath uses high-throughput microarrays to probe protein function.



Panomics' cytokine microarray.

to 48 protein microarrays in 1–2 hours at maximum capacity.

There has been considerable interest in recent years in applying proteomics to clinical diagnostics and predictive medicine. The goal is to identify disease markers, or biomarkers, that can be used to extract diagnostic (or even prognostic) information from body fluids such as serum, saliva or urine.

The SELDI-based ProteinChip System from Ciphergen Biosystems of Fremont, California, for example, can distinguish between patients with ovarian cancer and women without the disease. The approach seems promising as a means of early detection, but there are questions about the robustness and reproducibility of the serum proteomic patterns generated, which should be

answered after larger-scale clinical studies.

Serum is applied directly to the surface of a ProteinChip Array and proteins are selectively retained on a variety of chromatographic surfaces. After incubation, unbound proteins are washed off, an energy-absorbing matrix is applied, and the array is analysed by TOF-MS. Most recently, Ciphergen has changed the way it makes its arrays. "It's our goal to become a 'matrix-free' company," says Scott Weinberger, who runs Ciphergen's proteomics research group. Last month it launched a new line of Surface Enhanced Neat Desorption (SEND) ProteinChip Arrays, in which the matrix is incorporated in the surface chemistry of the array. This not only enhances throughput, but also improves sensitivity and sequence coverage. SEND ID, the company's first SEND product, is targeted at peptide mass fingerprinting applications.

To unravel the complexity and capture the dynamic nature of the human proteome, researchers will need to have all the tools in the proteomics tool-box at their disposal as no one technology will suffice. Making sense of the data will require sophisticated IT solutions to analyse, manage and exchange data and here proteomics lags some way behind genomics and gene-expression studies. But, ultimately, it will require the integration of proteomics data with data from gene expression studies and functional assays, as well as with structural information on proteins and protein complexes provided by X-ray crystallography and nuclear magnetic resonance spectroscopy. ■

Diane Gershon is assistant editor, *Nature Medicine* Technical Reports.