

Proteomics technology: Character references

technology feature

In the *Odyssey*, Homer's hero has his hands full when he faces Proteus. The demigod challenges Odysseus by transforming himself into a lion, a boar, a serpent, a wave and finally a tree. In proteomics, scientists trying to discern the nature of proteins face an equally formidable challenge, because protein data are as mutable as Proteus. Protein levels in different cell types change constantly as they are upregulated, downregulated, cleaved and phosphorylated.

Because protein information, unlike DNA, is not static in the cell, scientists must follow Odysseus' lead. They will have to be resourceful, especially as the tools used in today's high-throughput environment still bear the stamp of an earlier era when one protein at a time was the standard.

The 2D gel used to separate individual proteins from complex mixtures dates back to the mid-1970s. Mass spectrometry, which identifies proteins by weight once they are isolated, has been around since the First World War. And industrial robots,

used to usher the proteins through the intermediate steps that separate these two techniques, date back to the 1960s. Most venerable of all is the century-old separations technique of chromatography.

Fortunately for scientists aiming for widespread protein characterization in the wake of the triumph of genome sequencing, a series of improvements in mass spectrometry and 2D-gel technology is readying these tools for the task that lies ahead. Chromatography was modernized in the 1970s with the invention of high-pressure pumps, the addition of multiple columns and improved packing materials for columns, leading to its modern incarnation as high-performance liquid chromatography, or HPLC, a workhorse in many life-sciences labs.

Established scientific-equipment companies are also working to integrate more steps of the overall proteomics workflow into fewer pieces of equipment. And many start-up companies are looking for ways to enhance or supplant parts of the established proteomics process.



The mass spectrometer is key to proteomics.

SPL

Although there are many different methods emerging — from mapping all the proteins in a single organism to describing the multitude of interactions experienced by proteins during their lifespan — the general technique of isolating and identifying the many proteins in different cell types remains central.

MASS SPECTROMETRY

Mix and match

Mass spectrometry represents a worldwide market worth US\$1 billion a year, with about a third of that dedicated to machines especially suited for proteomics. The system uses three components — an ionization source, an analyser and a detector. Users have at least two choices for each component. Assorted pairings offer different advantages — some combinations are more suited

to proteomics, whereas others lend themselves more to small-molecule analysis. And some combinations will integrate with other proteomics equipment such as liquid chromatography.

Companies are tending to make their new machines more versatile, more automated and more compatible with other proteomics equipment — but, in general, the more choices offered by one

machine, the higher the price tag.

The choices begin where the process starts — ionization sources. Ionization gives the sample an electric charge. The widely used MALDI (matrix-assisted laser-desorption/ionization) uses solid samples, and produces ions of large and small molecules. Electrospray ionization (ESI) is used less often in proteomics. It ionizes liquid samples and is most often used for peptides and small molecules. It can be directly coupled to liquid chromatography systems.

For analysis, time-of-flight (TOF) is most frequently used with MALDI, whereas ESI is usually coupled to quadrupole or ion-trap analysers. Quadrupole machines are considered low-performance instruments compared with MALDI-TOF, but they only cost about a third as much. Ion-trap analysers are also modest performers, but they are robust and easier to look after than the other types, and are even more modestly priced.

Finally, there are two kinds of mass spectrometer — MS and MS/MS. MS is the faster, easier-to-operate option. But, in addition to generating a spectrum of the sample, MS/MS can take some of the ions that have been separated and measured, fragment them further, and then generate spectra of those parts. This allows users to discern which amino acids the peptides contain, and, in some cases, can identify the sequence of these amino acids within the peptide.



Integration: speeding analysis.

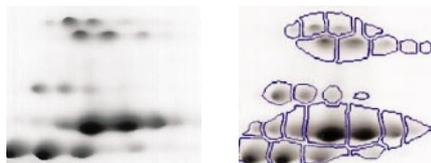
CYROS

There are several possible starting points for protein identification. But the most well-travelled route into proteomics starts with a sample in a 2D gel being fed into an electrophoresis machine. This is followed by either automatic or manual picking and excision of the protein spots of interest, which are then fed into a mass spectrometer (see 'Mix and match', previous page).

Celia Caulcott, who heads an effort by the UK's Biotechnology and Biological Sciences Research Council to develop new proteomics technologies, says that, despite a lot of R&D, traditional techniques for protein identification still stand. "The gels still seem to be the pre-eminent way people want to do things," she says. Beguiling techniques such as protein arrays, which could supplant gels if successful, have yet to prove they can be viable both scientifically and commercially, she says.

Joakim Rodin, director for proteomics R&D at Amersham Biosciences, a biotech-equipment company based in Uppsala, Sweden, agrees that the gel system, although not the easiest thing to work with, has yet to be supplanted. "It's still a lot of work running the gels," he says. But improvements in capacity, such as the company's Ettan Dalt II system, allows researchers to run up to 12 gels in parallel with more reproducibility and sensitivity.

And the gels themselves have improved,



Identifying spots on gels can be time consuming.

he says. They are getting bigger, so more sample can be loaded, which improves the detection of low-abundance proteins. 'Zoom' gels have also been developed with ever-narrowing pH ranges, which give better resolution as well as higher sensitivity.

Fluorescent labelling is also getting better, he says. Differential-expression analysis using difference gel electrophoresis, developed at Carnegie Mellon University, allows up to three samples to be run simultaneously on a single gel using cyanine-dye chemistry. This should let researchers detect protein differences between normal and cancerous tissues on the same gel. The method also allows multiplexing of gels, which significantly increases throughput, reproducibility and accuracy. Multiple gels provide comparative analysis and accurate measurement of differential protein expression. Although the handling and analysis of 2D gels have improved dramatically, Rodin notes that complementary techniques, such as X-ray

crystallography, are needed to resolve the whole proteome.

Fortunately, the next stage of the proteomics pipeline, handling the intermediate steps between electrophoresis and mass spectrometry, is becoming easier. Picking the protein spots off the gels, then digesting them into peptide fragments used to be two separate, manual tasks. Now they are becoming automated and are being integrated into the workflow (see 'Multiple choice', below). But improving and combining individual components can be challenging, says Steve Martin, director of Applied Biosystems' Proteomics Research Center in Framingham, Massachusetts. For example, increasing the capacity of one instrument without accounting for the additional need for throughput in others can actually result in bottlenecks, he says.

Three commercial — and by today's standards, integrated — systems are made by Amersham Biosciences, Genomic Solutions in Ann Arbor, Michigan, and Bio-Rad in Hercules, California. Their basic components are similar — they all use robotic sample-preparation, 2D-gel electrophoresis, excision of spots, labelling, and ionization and analysis of the peptide fragments by mass spectrometry. In these systems, data generated from all the instruments are presented in a user-friendly graphical interface.

AUTOMATION

Until recently, characterizing proteins was done slowly. But with the many candidates in the newly sequenced genomes crying out for attention, and the lure of complex protein assemblies beckoning, labs are gearing up to look at many proteins simultaneously.

The key to making such a system work lies in replacing error-prone humans with spot-picking robots, guided by cameras and sophisticated image-analysis software. The Australian company ARRM has a system that excises spots from gels or polyvinylidene fluoride membranes and places them in a 96-well plate for subsequent proteolysis.

Genetix, of New Milton, UK, uses a line of sample-preparation, gel-spotting and spot-excision units. Soon these will be joined by a machine to prepare MALDI samples automatically, thus helping to integrate raw samples and mass fingerprints. Genetix is also getting into chip arrays and yeast two-hybrid systems, two automated ways of looking at protein interactions. Other major players in lab automation are Amersham Biosciences in Uppsala,



Lab-on-a-CD systems from Gyros.

in Uppsala. Gyros has updated an idea from the 1970s by engraving microscale channels and mixing chambers on a compact disc. Centrifugal force and controlled surface chemistry are used to regulate the flow of liquid through the CD. Despite the small size of the system, the price tag will probably ensure that it will mainly be used by big pharmaceutical companies or 'protein factories' rather than small independent labs.

Multiple choice

Sweden, Bio-Rad in Hercules, California, and Genomic Solutions in Ann Arbor, Michigan.

Large Scale Biology in Germantown, Maryland, and Oxford Glycoscience in Cambridge, UK, aim to automate the entire protein-discovery process in humidity-controlled, robot-populated buildings. Here massive amounts of samples would travel through the pipeline from gel to mass spectrometer and data.

But harking back to the idea that small is beautiful, another school of thought sees a nano future for the science in 'lab-on-a-chip' technologies such as those of Caliper, of Fremont, California, and Gyros

These stations are quite expensive but, just as core facilities for genome sequencing sprang up once the equipment came of age, the same is likely to happen with protein characterization. This should ensure that smaller academic and commercial labs will share in the advance of knowledge. And smaller labs might still be able to automate individual steps, such as spot picking or digestion, finding new ways to integrate steps that might be overlooked in larger, more streamlined organizations.

Alternatives for eliminating, rather than integrating, such steps are also emerging. One fairly new strategy involves transferring the gel to a membrane made of polyvinylidene difluoride (PVDF), then probing the membrane directly with mass spectrometry. This bypasses the spot-cutting step between electrophoresis and mass spectrometry.

Improvements also extend to mundane but essential items such as stains. Coomassie blue, a staple in most labs, can interfere with the digestion of gel spots by trypsin, so new stains such as zinc imidazole and noncovalent fluorescent SYPRO dyes, which do not have this limitation, are being introduced.

Mass-spectrometry output

It was not until the early 1990s that the mass spectrometers, now virtually essential components in the proteomics pipeline,

could be used to analyse proteins.

Mass spectrometry relies on the fact that a substance carrying a net electric charge — an ion — can be made to move in a predictable way in an electromagnetic field. Ions are sorted by their charge-to-mass ratio, and from these a 'mass fingerprint' of the sample can be derived. Software, such as the University of California's Prospector package, can then be used to match the fingerprint to a protein database such as Amos Bairoch's Swiss-Prot (see 'Setting standards', overleaf).

In earlier models, excessive ionization energies would blast delicate molecules such as DNA and proteins into indecipherable particles. But innovations using a matrix such as MALDI, which protects the sample by modulating the ionizing laser beam, have helped to overcome this limitation.

Nevertheless, the technique still has its limits. A mass fingerprint will not be enough for identification if the protein is not registered in a database, or if post-translational modifications have changed its observed mass from the predicted value. In these instances, more information can be obtained from secondary protein fragments by re-routing the ions from the first analysis down a second channel and then analysing these fragments with the spectrometer. Of course, more complete databases will also help. And pairing mass

spectrometry with other techniques, such as some kinds of protein-detector chip (see 'Alternative approaches', below) may make the method even more useful.

Future challenges

Automating and integrating the protein-characterization process is a good start, but there is no simple way forward. Although effective with adequate sample sizes, automated processes in general are not effective with very small amounts (less than 10 femtomoles of material).

It is hard enough to describe a single protein in a particular state. But things get even more difficult when trying to



Leroy Hood (right) and Ruedi Aebersold.

CHIPS

Proteins lack DNA's copying ability and do not readily undergo amplification, making separation and fractionation more important — especially for small amounts of proteins. And the inherent complexity and diversity of proteins makes a viable protein array an even more difficult goal. But the need to process proteins en masse is so urgent that heroic efforts are under way to develop a workable protein chip.

Leading the field at present are designs based on antibodies tethered to a solid surface. Large Scale Biology in Germantown, Maryland, and Biosite Diagnostics in San Diego, California, are developing an array of antibodies against 2,000–5,000 protein targets from the former's human protein index database. Biosite will use its omniconal phage display technology to generate high-affinity antibodies against the targets. The companies hope the system will be available in the second half of 2002.

But an inherent drawback of antibody chips — or any protein chip, for that matter — is the destructive effect of proteases that may be lurking in the analyte mixture. "You have to use protease inhibitors if you're sampling microdissected tissue," says Lance Liotta of the US National Cancer Institute's Center for Cancer Research, who invents tools for proteomics and has surveyed the existing technology. "Process the tissue, lyse it, stain it and pray

Alternative approaches

that these manipulations don't affect the 3D state of the protein."

Perhaps the biggest challenge is the accurate quantification of low-abundance protein. The faint signal of a protein of interest may easily be swamped by the much higher concentrations of other surrounding proteins.

Ciphergen in Fremont, California, is selling a device that helps scientists to detect low-abundance proteins. The company's chip uses specific surface chemistries to affinity-capture minute quantities of proteins. "A peak in one sample but not the other says a variation exists, but you still have to figure out what it is," says Mike Baldwin, a chemist at the University of California, San Francisco. "It's an interesting approach, but not mainstream proteomics — at least, not yet."

Another recent quantitative protein-expression and -identification technique using mass spectrometry is isotope-coded affinity tagging (ICAT), a kind of labelling invented by Ruedi Aebersold at the Institute for Systems Biology in Seattle. The start-up company Sense Proteomic, based in Cambridge, UK, is trying to use smaller numbers of mounted proteins to assay for suspected protein-protein interactions such as those known to play a role in toxicity.

Other chip approaches towards proteomics include atomic-force microscopy, aptamer libraries and biosensors.

characterize thousands of proteins active at any time in various parts of the cell. Michael Washburn and Dirk Wolters at Syngenta Agricultural Discovery Institute in San Diego and John Yates at the Scripps Research Institute in La Jolla, California, have devised a system to separate and identify 1,484 proteins from the proteome of the yeast *Saccharomyces cerevisiae* (see *Nature Biotechnol.* **19**, 242–247; 2001). But that relatively low number in the humble yeast doesn't begin to reveal the complexity in humans. For example, there are a thousand or more proteins involved in the G-signalling pathway, which regulates everything from the most basic activities of the cell (division, motility) to the most specialized

ones (secretion, electrical excitability).

Perhaps the biggest hurdle is not in designing the equipment but in the conceptual realm. Researchers might know individual elements in a signal cascade, understand something about their function, and perhaps even have obtained their structure. But, explains Ehud Isacoff, a biophysicist at the University of California, Berkeley, scientists are still encumbered by a bias to view the overall picture as if it were made up of discrete events, with one protein handing a signal to another sequentially, in a series of 'stills'.

What is really happening in the cell, Isacoff continues, "is that proteins are very localized, and dock against one another very precisely in assemblies, and signalling happens by molecular motions that propagate from one subunit to another". New methodologies and systems of notation must be devised to describe these things, and a new breed of student has to be recruited who can think about them as concrete objects with specific structures and interactions.

In fact, these needs are being recognized and the integrative effort is under way on several fronts. Leroy Hood's Institute for Systems Biology in Seattle has been in existence since early last year (see *Nature* **407**, 828–829; 2000), and Al Gilman's Alliance for Cell Signalling at Dallas set up shop a year ago (see *Nature* **407**, 7; 2000).

They aim at a holistic understanding of the cell in all of its pathways and interactions. New methodology — and, perhaps, improved equipment — may emerge from such efforts.

And a Clinical Proteomics Initiative, under the aegis of the US National Institutes of Health, started seeking grant applications last month. One of its key elements will be the antibody consortium, says Lance Liotta of the National Cancer Institute and one of those engineering the enterprise. This will be modelled on the open-access but industry-supported SNP consortium that is mapping simple genetic variations. Support — both in terms of finance and willingness to donate antibodies — from industrial and academic groups is very enthusiastic, says Liotta. The consortium's ultimate goal is to develop and make available arrays of every antibody and every ligand in existence.

Other aspects of the NIH initiative are looking for new approaches to existing techniques. However, it's unlikely that any new technology will completely replace an old one. Instead, innovations arising from the initiative will probably occur alongside the stalwarts of electrophoresis, mass spectrometry and chromatography — further complicating the ever-changing face of proteomics. ■

Potter Wickware is a science writer based in San Francisco; Paul Smaglik is *NatureJobs* editor.



Al Gilman: seeking the cell's secrets.

SOFTWARE

In the realm of software and databases, there is a real opportunity for integration, but instead developers have tended to go off in their own directions. Great strides have been made in areas such as image analysis and peak-picking tools for mass spectrometry with software packages including Tycho, Melanie and Quest. Software developed by Nonlinear Dynamics of Newcastle upon Tyne, UK, aids in spot detection on gels and also helps in quantitative analysis of those spots once they are picked. Major equipment manufacturers



Tony Pawson.

Amersham Biosciences and PerkinElmer have already signed on to bundle this program, called Progenesis, with some of their instruments. But, according to Patsy Babbitt, a protein informaticist at the University of California, San Francisco, the software side is fragmented. "It's a big problem," Babbitt says.

Organizations such as the Bio-Ontologies Consortium aim to clarify the picture with

standards and nomenclature, but perhaps what is lacking are new ways of thinking about the information generated in proteomics — classical bioinformatics is built around pattern-matching algorithms.

Tony Pawson and Chris Hogue at the University of Toronto have been thinking about the informatics side of proteomics. They have developed the Biomolecular Interaction Network Database (BIND), which indexes interactions between DNA, RNA, proteins and small molecules, as well as temporal and compartmental information. As BIND's content grows, "we'll be the GenBank of interactions", predicts Francis Ouellette, of the University of British Columbia in Vancouver, one of the resource's developers. Other databases for proteomics include the Database of Interacting Proteins at the University of California, Los Angeles, Large Scale Biology's Human Protein Index and Atlas Base, by the San Diego company Accelrys, which contains protein structures.



Chris Hogue.

Setting standards