A post-genomic challenge: learning to read patterns of protein synthesis

The new science of proteomics promises exciting insights into the working of the cell. But major technical hurdles remain to be overcome.

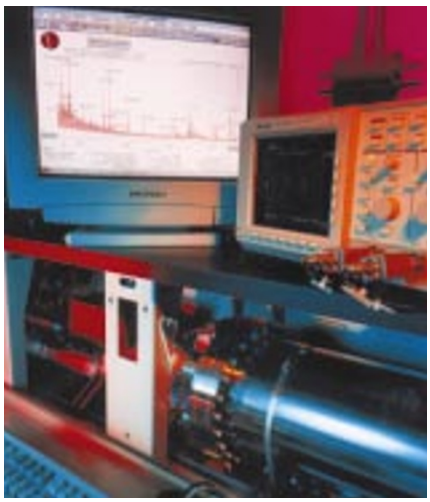
As the full-length sequences of target genomes come into view, biologists know they represent just the start of a long march towards an understanding of how organisms, including humans, develop and function. To many, the next key landmark will be an overview of the characteristics and activity of every protein that an organism can synthesize in its lifetime: its 'proteome'.

Because the technology required is still at a rudimentary level of development, and is more troublesome than genomic technologies, completion of such a task cannot be expected for many years. But that is not stopping biologists from setting out along the way.

Ten years ago, a protein chemist would have been happy to identify two or three proteins a year. Now, increasing volumes of genome data, combined with mass spectrometry technologies, permit a researcher to identify hundreds of proteins within a week.

Although the technologies that have revolutionized protein chemistry are in their infancy, some scientists are already starting to cautiously discuss the concept of a Human Proteome Project, to complement the Human Genome Project. Others say it is far too early to speak so ambitiously, or suggest that the idea of adopting the systematic approach of the genome project is not sensible, given the ever-shifting populations of proteins.

But the idea is seductive. By identifying proteins on the scale of the proteome — which can involve tens or even hundreds of thousands of proteins, depending on the state of the cells being analysed — proteomics can answer fundamental questions about biological mechanisms at a much faster pace than the single-protein approach.



Fast fingerprinting: highly automated mass spectrometry identifies proteins rapidly.

The 'global picture' painted by proteomics can, for example, allow cell biologists to start building a complex map of cell function by discovering how changes in one signalling pathway — the cascade of molecular events sparked by a signal such as a hormone or neurotransmitter — affect other pathways, or how proteins within one signalling pathway interact with each other. The 'global picture' also allows medical researchers to look at the multiplicity of factors involved in diseases, very few of which are caused by a single gene.

Such understanding can be exploited to identify better drug targets — proteins with which drugs selectively interact to achieve a defined therapeutic response. It also offers drug companies the prospect of being able to rapidly identify the ultimate number of useful protein drug targets — reckoned to lie between 1,000 and 4,000 — and to identify individuals destined by their proteomes to suffer side effects from drug therapy.

However, major technical problems need to be resolved before proteomics can become the large-scale, highly automated affair that genomics has become. There are two principal steps in proteomics (see page 716). The first is separating proteins in a sample, usually using two-dimensional (2D) gels which separate according to molecular charge in one direction and mass in a perpendicular direction. The second is identifying the separated proteins, typically using new mass spectrometry techniques and bioinformatics. Each step has limitations that make automation difficult, but the strongest curses of proteomics researchers are reserved for the 20-year-old 2D gel

OGS

Proteomics, transcriptomics: what's in a name?

The term proteome, coined in 1994 as a linguistic equivalent to the concept of genome, is used to describe the complete set of proteins that is expressed, and modified following expression, by the entire genome in the lifetime of a cell. It is also used in a less universal sense to describe the complement of proteins expressed by a cell at any one time.

Proteomics refers to the study of the proteome using technologies of large-scale protein separation and identification. The nomenclature is catching on. The generation of messenger RNA expression profiles is referred to as transcriptomics, as these are based around the process of transcription. And the complement of mRNAs transcribed from a cell's genome is called the transcriptome.

The company Ciphergen, based in Palo Alto, California, is trying to popularize the term phenomics to describe the technology of automated functional analysis of proteins. The word derives from phenotype — the observable characteristics conferred by a gene. Some enthusiastic researchers in the field are even starting to refer to the whole operation of molecular analysis of a cell, extending from DNA through RNA to protein, as operomics.

© 1999 Macmillan Magazines Ltd

separation technology. Only the truly green-fingered can reproducibly separate a complex solution of proteins on a gel matrix.

Despite much refinement, it is still referred to as 'the black art' and 'the dark force of proteomics' — but 'a necessary evil', which requires skilled staff. There is a limit to how many proteins a single gel can separate, and the sensitivity is still not adequate to detect proteins appearing at very low levels. Some classes of protein, particularly hydrophobic membrane-bound proteins that are favourite targets for drug development, will not run on 2D gels. On the other hand, the 2D gel is the only method that can resolve large numbers of proteins in a quantitative way.

After the genome

Now that the sequences of genomes from several species have been, or are about to be, completed, researchers and their funding bodies — both public and private — are looking towards the next step: the understanding of gene function. Of the roughly 140,000 genes in the human genome sequence, which will be completed within four years, the function of an estimated two thirds is likely to be unknown.

Protagonists argue that proteomics is one of the most important of the so-called 'post-genomic' approaches to understanding gene function because it is the proteins expressed by genes that are ultimately responsible for all processes that take place within the cell. But, while proteins may yield the most important clues to cellular function, they are also the most difficult of the cell's components to detect on a large scale.

A second, complementary, post-genomic approach is expression profiling, also known as transcriptomics. When a gene is expressed in a cell, its code is first transcribed to an intermediary 'messenger RNA' (mRNA) which is then translated into a protein. Transcriptomics involves identifying the mRNAs expressed by the genome at a given time. This gives a snapshot of the genome's plans for protein synthesis under the cellular conditions at that moment.

Transcriptomics has the advantage over proteomics that the technology is simple and lends itself readily to automation and high throughput. Messenger RNAs can be fished out of a cellular soup onto microarrays. These are chips onto which are stuck thousands of complementary DNAs derived from as many genes in a genome as have been sequenced, or are deemed necessary for a particular experiment. Each cDNA specifically binds the mRNA transcribed from the gene from which the cDNA is derived. If only a small amount of mRNA is present, this can be 'amplified' by polymerase chain reaction techniques.

Transcriptomics can therefore yield important biological information about what genes are turned on, and when. But it has the disadvantage that, although the snapshot it provides reflects the genome's plans for protein synthesis, it does not represent the realization of those plans.

The correlation between mRNA and protein levels is poor, generally lower than 0.5, because the rates of degradation of individual mRNAs and proteins differ, and because many proteins are modified after they have been translated, so that one mRNA can give rise to more than one protein. Even in the simplest self-replicating organism, *Mycoplasma genitalium*, there are 24 per cent more proteins than genes, and in humans there could be at least three times more. Post-translational modification of proteins is important for biological processes, particularly in the propagation of cellular signals, where, for example, the attachment of a phosphate group to a protein can trigger either activation or inactivation of a signalling cascade.

So measuring proteins directly gives a more accurate picture of a cell's biology and, if it were as technically easy as measuring mRNAs, more scientists would choose to do so. Ian Humphery-Smith is one scientist who believes that technical difficulties should not stand in the way of a systematic effort to analyse the whole human proteome. His former group, from the Centre for Proteomic Research and Gene Product Mapping at the University of Sydney, has published the most complete proteome to date — almost three-quarters of the predicted proteome of *M. genitalium*, whose relatively small genome (less than 500 genes) has been fully sequenced.

"It's small, and it is not complete, but three quarters of everything is very relevant", says Humphery-Smith, who has recently taken up the new chair of pharmaceutical proteomics at the University of Utrecht in the Netherlands. Proteomic researchers are at the same scientific and technological stage as genome researchers were in 1986 when the Human Genome Project was first seriously proposed, he says, "so the time is right to launch a Human Proteome Project now".

This project should identify, catalogue and annotate the entire human proteome — the expression of every gene: "Now we can do simple prokaryotes, we should start systematically to analyse the proteomes of lower eukaryotes whose genomes are near completion, so that we will be ready to approach the

How to spot a protein in a crowd

The two key steps in classical proteomics are the separation of proteins in a sample derived from cells or tissues, and their subsequent identification. The best separation method is 2D gel electrophoresis, in which spots of a carefully prepared mixture of proteins extracted from cells or tissues are applied to a polyacrylamide gel.

The proteins, which can number tens of thousands, are separated along the gel in one direction according to their molecular charge, by applying an electric field. This process is called isoelectric focusing.

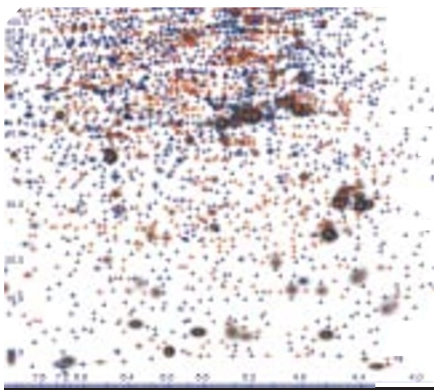
They are then separated in another direction on the basis of their molecular mass, using SDS-PAGE (polyacrylamide gel) electrophoresis, a molecular sieving method. A typical gel can reliably separate 2,000 protein spots in this way. The 'best' gels can separate up to 11,000 protein spots.

Proteins separated on the gel are stained using the Coomassie blue dye, silver stains, fluorescent dyes or by radiolabelling, and then quantified using spectroscopic or radiographic techniques. As yet, amounts of proteins of less than one nanogram cannot be routinely measured.

Although 2D gel electrophoresis gives the highest resolution of all available methods, it is notoriously difficult to carry out, and it cannot detect some of the most interesting proteins in a cell.

Hydrophobic proteins, for example, such as the cell-membrane-spanning receptor proteins that are such attractive drug targets, simply do not dissolve in solvents used for isoelectric focusing. Neither do proteins with very high relative molecular mass. Low-abundance proteins, which tend to do the most interesting jobs in a cell, are drowned out by high-abundance 'housekeeping' proteins, which can be present at 10,000 times the concentration.

Furthermore, there is no amplification method for proteins, analogous to the



DANISH CENTRE FOR HUMAN GENOME RESEARCH

Separated out: a 2D gel displays more than 2,000 proteins, known (red) and unknown (blue).

© 1999 Macmillan Magazines Ltd

human proteome when the genome is completed in 2003. This would allow essential infrastructure to be put in place that could also be used for the human proteome."

Mathias Uhlén, director of the new Centre for Proteomics at the Royal Institute of Technology in Stockholm, also believes it is "time to talk about a global proteomics project — and to set endpoints for the project which obviously cannot be as simple as the single endpoint of the Human Genome Project".

And Hans Lehrach, a director of the Max Planck Institute for Molecular Genetics in Berlin, and spokesman for the German Human Genome Project, sees the systematic cataloguing of the proteome as almost a duty for the genomics community, to provide full information for biological researchers.

This view is not widely shared among the as yet small international community of proteomic researchers. Most feel that the technical hurdles are still too high to launch such a major initiative in the near future. But Humphery-Smith argues that serious investment in a Human Proteome Project would rapidly drive the development of technologies to complement the limited 2D gel approach.

Craig Venter, the genome technology guru whose company Celera Genomics aims to sequence the human genome by 2001, is convinced that this is the right way for scientists to work. "Scientists must work with the tools at hand," he says. "It would have been wrong for molecular biologists to have waited until the sort of high-throughput technology we now have at Celera became available before launching the Human Genome Project."

But Leigh Anderson, director of one of the world's two largest proteomics companies, Large Scale Biology Corp., which works

Proteomics is one of the most important of the 'post-genomic' approaches to understanding gene function.

with most of the world's top pharmaceutical companies on specific proteomics projects, is one scientist who is unenthusiastic about a Human Proteome Project. He actually proposed such a project, then called the Human Protein Index Project, nearly 20 years ago, which was discussed at high political levels before being superseded by the genome revolution. "But we were naïve then," he says.

He now thinks that "large-scale proteome research should be carried out in companies rather than through government-sponsored projects". Unlike the genome, which is a constant entity, interesting biological and medical questions relate to the proteome expressed in different tissues, different disease states and under different conditions, he says: "This prospect of immediate application is why proteomics is so commercially interesting at the moment".

"Enthusiasm, scepticism and indifference: this is the same range of opinions that were expressed when the Human Genome Project was launched," says Mike Ashburner, co-director of the European Bioinformatics Institute at the Sanger Centre in Cambridge, England. Scientists were daunted by the size of the task of sequencing genomes from whole organisms, he says, and many questioned whether the investment required was justified. But experience bears out the wisdom of taking the plunge.

Clinical applications

While some scientists talk in terms of ambitious 'cataloguing' plans, pharmaceutical companies are already working on the application of proteomics in disease, hoping to identify new diagnostic tools or leads for the development of drugs.

Identifying disease markers, proteins that appear or disappear during the course of a disease, does not necessarily require that all expressed proteins in a clinical sample be identified — although the more complete the proteome, the more complete will be any set of markers. Service companies such as Large Scale Biology Corp. and its UK competitor Oxford GlycoSciences (OGS), which have developed automated and high-throughput proteomics technologies, can identify high proportions of the total protein expressed in a tissue with reasonable efficiency.

For example, under a contract with Pfizer, OGS has analysed hundreds of samples of cerebrospinal fluid from patients in different stages of Alzheimer's disease, collected by Pfizer and the National Institute of Mental Health, a division of the US National Institutes of Health (NIH). Results have not yet been formally announced. But key proteins have been identified whose expression changes during the progression of the disease.

Scientists in such big companies tend to be scathing about the attempts of small academic groups to set up similar clinical research projects with more modest

Current methods separate proteins on gel, then identify them using lasers.

polymerase chain reaction method for amplifying genes, to help make up for this difference in abundance. And proteins with very high molecular charges or very low molecular mass will not separate on gels.

More proteins can be separated by using time-consuming tricks such as fractionating cells into their various organelles before running consecutive gels, or sequentially running 'zoom' gels, which separate proteins according to a series of narrow isoelectric ranges. But, even after such tricks have been exhausted, around 20 per cent of the proteome may still be missing.

Once proteins have been separated, visualized and quantified, they must be identified. Spots are excised from the gel — companies such as Oxford GlycoSciences and Large Scale Biology have robots to do this — and proteins are digested into fragments by specific proteases such as trypsin. The fragments are analysed by mass spectroscopy in a process called peptide mass fingerprinting, in which proteins are identified by comparing the mass of the peptide fragments with data predicted by genetic or protein sequence information.

The most commonly used type of mass spectrometer is called MALDI-TOF, in which protein fragments in a solid-phase sample are ionized by a laser beam. The MALDI-TOF is highly automated, and popular with academic research groups.

A more sophisticated, but more time-consuming, method is tandem mass spectrometry. Each peptide analysed by mass spectrometry is subjected to further fragmentation and mass spectrometry, to give partial information about the peptide sequence. This information is helpful if full genomic data are missing, because proteins can be identified from databases of expressed sequence tags, which have more entries than standard genomic databases.

Any experiment that involves a limited number of proteins, such as those in a purified protein complex, can avoid using 2D gel separation, which is tedious and difficult, by using methods such as high-performance liquid chromatography, gel filtration or one-dimensional gel chromatography.

These methods separate and concentrate proteins in a complex sufficiently well to be identified by mass spectrometry, although they do not allow quantification. The ultimate goal, of course, is to be able to avoid having to separate proteins at all. The protein sample would merely be run over arrays impregnated with antibodies, or other types of affinity probe, to identify all proteins of interest — this is the promise of chip technology **(see page 718).** **A. A.**

© 1999 Macmillan Magazines Ltd

equipment. Chris Ashton, until recently with OGS, doubts that small units without access to sophisticated robotics to allow high throughput should get involved. "Proteomics is a big science which is technically much more complicated than genomics, and so best handled by big companies or their equivalent," he says.

But Peter Mose Larsen, director of the Centre for Proteome Analysis at Odense University in Denmark, is one of many academics who disagree. His lab has identified and patented two groups of protein markers of hypertension which are being tested by five pharmaceutical companies as diagnostic tools, and in some cases as drug targets.

He admits that inexperienced groups often underestimate the skill required in running 2D gels. But lots of small-scale science is required, which does not need the highest proportion of expressed proteins to be detected, nor the most efficient robotics, he says. "The contribution of many small groups is vital; and academic groups can work as well as large companies, with equal success." His centre is one of ten partners in a European Commission initiative to develop generic proteomics technologies and make them available to smaller laboratories, and also small companies, throughout Europe.

In fact, small — and large — academic research units springing up around the world are starting to identify markers in a range of conditions, from organ rejection to cancers, and a wave of publications is likely to hit the literature soon.

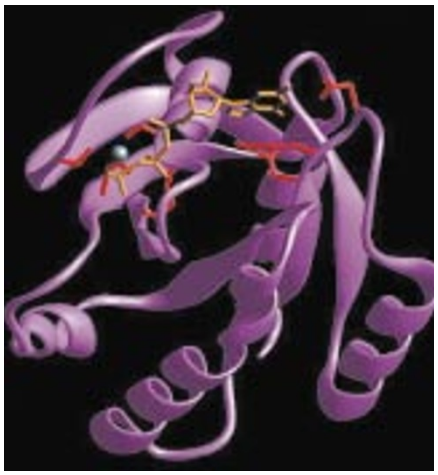
Two centres of the Ludwig Institute for Cancer Research, for example, in London and Melbourne, are running programmes to determine protein expression in breast and colon cancers. Director Mike Waterfield says these programmes, carried out in collaboration with OGS, "will complement the transcriptomics and genomics efforts that the Ludwig Institute is carrying out with the Imperial Cancer Research Fund and the Sanger Centre near Cambridge". This collating of information at all three molecular levels, sometimes jokingly referred to as 'operomics', is an increasingly frequent feature of research programmes.

The National Cancer Institute (NCI), a division of the NIH, funds similar research through an extramural programme concerned with developing and using molecular-based technologies, including proteomics and transcriptomics, to aid tumour classification.

One of the biggest centres to be funded is that of Samir Hanash, from the University of Michigan School of Medicine, who has a new $10 million NCI grant to generate proteomic data on lung, colon, liver, breast and ovary cancers, which will complement genomic and transcriptomic data from the same tissues, and create a database combining all the information, including original 2D gel images.

The NCI is also co-sponsoring, with the Food and Drug Administration (FDA), a new multi-million dollar Tissue Proteomics Initiative run by Lance Liotta, a principal investigator at the NCI laboratories in Bethesda, Maryland, and molecular biologist Emanuel Petricoin, a senior staff fellow at the FDA. This will generate proteomic data in a range of cancers at different stages of disease and therapy — as well as developing proteomic technologies suitable for clinical work.

The FDA is interested in particular in finding surrogate markers of toxicity and measures of the efficacy of both new and existing drugs. In one project, Petricoin is monitoring changes in protein profiles, and some signal transduction pathways, in breast cancer epithelium and in body fluids before and after treatment with herceptin, a humanized monoclonal antibody that was recently approved by the FDA for late-stage breast cancer treatment. The changes will be correlated with side effects and clinical efficacy.



CORBIS

One among many: the p-21 H-ras protein, which is implicated in cell-signalling functions.

Petricoin says the programme aims to demonstrate 'proof of concept' to drug companies that remain uncertain about the impact proteomics is likely to have on their drug development programmes.

Petricoin is convinced the impact will be large, allowing, for example, companies to predict the toxicity of a potential drug much earlier than at present by looking for surrogate markers in the tissues of healthy volunteers in phase one clinical trials. At present, and in the absence of surrogates, toxicity often only becomes apparent when a drug is used in

Can researchers find recipe for proteins and chips?

Automated chip-based technologies for analysing thousands of proteins simultaneously, analogous to the cDNA chip-based technologies that have facilitated transcriptomics, could provide a leap forward for proteomics research, whose progress is limited by the cumbersome multi-step methods currently available.

Since the chemistry of proteins is orders of magnitude more complicated than DNA chemistry, the leap forward is unlikely to be seen in the next few years. Despite this, scientists are already starting to work on the problem, using different approaches.

The Stanford University laboratory of Patrick Brown, for example, renowned for its success in developing cDNA microarrays for high-throughput analysis of gene expression, is investigating whether similar success can be achieved with proteomics.

CARINA DENNIS

Brown: success on microarrays.

Neither Brown, a paediatrician with a doctorate in biochemistry, nor his postdoc, chemist Brian Haab who is working on the project, are daunted by the difficulty of the task. "Everyone recognizes it is worth doing," says Brown.

Haab hopes to become the next young star from Brown's lab, which has already produced Dari Shalon, inventor of a robotic arm that is an important component of microarray analysing systems, and Joseph DeRisi, famous for producing elegant experimental results of gene expressions.

Funded partly by the US National Cancer Institute, Brown's proteomics project has completed a pilot study designed to demonstrate the feasibility of microarray analysis of proteins. Antibodies are attached to microarrays, and these bind to and identify proteins that have been tagged with fluorescent dyes to allow quantification. Haab has already been able to correctly identify 40 known proteins.

Researchers in Brown's laboratory hope to compare hundreds or thousands of protein expressions to identify a pattern of variation from human sera or fluid samples, as is now being done with 2D gel technology — but much faster.

Some remain sceptical whether the microarray concept will work broadly enough on proteins, whose diverse chemistry poses serious difficulties, to be helpful in large-scale expression proteomics where thousands of unknown proteins must

© 1999 Macmillan Magazines Ltd

large patient cohorts in phase three clinical trials, when tens of millions of dollars may already have been spent unnecessarily.

Also, he says, drug companies could provide information on specific protein markers of toxicity or responsiveness to clinicians who could then screen out patients likely to be sensitive to toxic effects, or unresponsive to a drug, and also monitor patients for early signs of toxicity and efficacy.

Petricoin says that such screening could be done without using 2D gels. Technologies such as SELDI and liquid chromatography and mass spectrometry (see page 716), which are easier to automate, would suffice because the relevant proteins would be known, and would be limited .

Basic biology

These simpler proteomics techniques, which sidestep the 2D gel problem, are also being used by basic biologists to address fundamental questions, such as how proteins in a cell interact with each other in response to internal and external signals. This approach represents a step in the direction of creating a 'virtual cell' which will allow biologists to approach the ultimate question: how do living organisms operate as systems at the molecular level? (See also *Nature* **402,** 219; 1999.) The number of proteins involved in a typical pathway of protein interactions is usually less than 100.

Signalling pathways, the chain of protein interactions and other molecular events that follow activation of a receptor protein by a hormone or neurotransmitter, are a fundamental part of this activity, and a major new focus of research. For example, Tony Pawson, head of the programme in molecular biology and cancer at the Samuel Lunenfeld Research Institute at the Mount Sinai Hospital in Toronto, recently received a Can$12 million (US$8 million) grant to study signalling networks on a large scale.

Another key player is Matthias Mann at the University of Southern Denmark in Odense. Mann recently identified the 70-odd components of the spliceosome, a complex of proteins in all cells that act together to ensure that non-coding regions are cut out of mRNA molecules before these mRNA proteins are translated into proteins. Mann pioneered the 'affinity probe' approach to analysing protein interactions. This involves taking one protein, or a part of a protein that has an important binding domain, and simply seeing what sticks to it.

Another approach to analysing protein interactions is the so-called yeast two-hybrid approach, developed by Stan Fields at the University of Washington in Seattle, which uses the yeast genome as a sort of matrix for studying interactions between two proteins which can be from any species.

Hybrids are made of two proteins whose potential interaction is to be analysed: one test protein is attached to a yeast DNA-binding protein, and a second to the so-called transactivation domain, or gene activator, of a yeast transcription factor. If the two test proteins interact, a 'reporter' gene, typically a gene that turns yeast blue, is activated by the transactivation factor which has been pulled into the gene's vicinity on the yeast genome by the DNA-binding protein. This is a potentially powerful technique for large-scale studies.

> The prospect of immediate application is why proteomics is so commercially interesting now.

Pawson says that biologists, who have traditionally devoted their careers to the study of one protein in one pathway to broaden their perspective, are now realizing that more can be gained by using the new technologies to look at all proteins in a pathway at once. It is important, he says, that biologists work with scientists from other disciplines, such as informatics and engineering, to ensure that the right technology is developed. "Like genomics, proteomics biology is becoming technology-led."

Pawson is the first to agree, however, that technology is necessary but not sufficient for the advancement of proteomics. Julio Celis, director of the Danish Centre for Human Genome Research, has a word of caution for starry-eyed newcomers to proteomics: "New technologies are not a substitute for scientific rigour." Celis is often called the founding father of proteomics because he was the first to develop methods to identify multiple proteins separated by 2D gels in the early 1980s.

"Whoever manages to master the biology of the system they are working on will dominate the proteomics race," he says. Just seeing the level of expression of a protein rise or fall in a tissue sample may not be biologically meaningful. "You need to know if the protein in question has been transferred to a different cellular compartment, if it has disappeared from one cell type only to appear in a neighbouring cell type, or whatever."

Chip-based technology could speed the pace of protein identification.

Aebersold: 'take a global view'.

be identified. DNA is static, the protein complement of a cell is constantly changing, and proteins fold in many known and unknown ways that affect their function.

"It will be very difficult to find binders, or antibodies, for every single protein, and then detect what is bound to each in a quantitative way," says Ruedi Aebersold of the department of molecular biotechnology at the University of Washington in Seattle.

"If this is to be a global effort, you have to have conditions that will work for all proteins," says Aebersold. "There are more straightforward ways to deal with proteins than what they are doing in the Brown lab."

But Aebersold agrees with the Brown lab's approach to using protein arrays for disease diagnosis. "That type of assay has a lot of potential in diagnostics, where you look for a particular constellation of proteins. You optimize conditions for specific proteins; you know what you are looking for."

Aebersold's own lab is developing a different 'chip' approach, which involves miniaturizing classical electrophoresis technologies in a way that is compatible with automation. Proteins are separated on channels scratched onto chips by applying an electric field. They are then sprayed directly from the chip into a mass spectrometer for identification.

According to Ian Humphery-Smith, of the University of Utrecht in the Netherlands, both methods suffer the intrinsic problem of miniaturization — the small volume of protein solution that can be applied to the surface of a chip, which limits the detection of low levels of proteins. Detection of low-abundance proteins is a problem inherent to all proteomics technologies, but it is exaggerated as analysis systems are reduced in size. "We will need a major technological advance in protein detection systems for chip technology to move forward."

A second "absolutely necessary" technological advance required for the Brown approach, he says, is in the area of antibody production, which is slow and cumbersome. Scientists are already tackling this problem (see page 720).

It may take many years. But if a Human Proteome Project is to be achieved, then array-based procedures of one sort or another will be needed to cope with the enormity of the task. **Rex Dalton & A. A.**

© 1999 Macmillan Magazines Ltd

His own clinical research programme in bladder cancer not only generates classical mRNA expression and proteomic data from normal and diseased biopsy samples, but also identifies the exact cellular location of proteins in the same biopsies using specific antibodies. This is time-consuming, particularly as most of the antibodies required have to be made in-house. Classical production of monoclonal antibodies in mice takes many weeks.

Many biologists believe that finding new ways to make high-quality antibodies on a large scale could lead to a second proteomics revolution. Promising ideas for efficient antibody generation are already being tested. One of these, pioneered by Greg Winter at the Laboratory of Molecular Biology in Cambridge and the company Cambridge Antibody Technologies, uses bacterial phages to generate antibodies. Another idea, pioneered by the Swedish Centre for Proteomics in Stockholm, uses combinatorial protein chemistry to generate potential artificial antibodies.

If these, or any other method, could lead rapidly to the creation of a library of highly specific, high-affinity antibodies that could reliably identify each individual protein in the proteome, proteomics could become a highly automated, high-throughput science like genomics. The antibodies could eventually be placed on chips, much like the cDNA arrays used in transcriptomics (see page 718).

Even in the near future, however, proteomics conducted on a limited scale has much to offer clinicians and basic biologists, and the decisions of funding agencies show that this is becoming more widely recognized (see below). Indeed, these decisions suggest that it may not be premature to start to discuss the realization of a longer-term goal, a Human Proteome Project.

Such a move would provide an invaluable tool for basic biologists and pharmaceutical companies alike, just as the Human Genome Project is providing a freely available tool in the form of complete genomic information. It would be several orders of magnitude more complex. But it would help drive the development of technologies to speed up the task, leading to a far more complete understanding of the operation of the human body than genomics alone can ever provide. **Alison Abbott**

Funding agencies move hesitantly to embrace novel approaches

One of the first funding agencies to recognize the potential of protein analysis was the US National Science Foundation. In 1989 it agreed to support the start of a ten-year programme at the University of Washington in Seattle to create a centre in molecular biotechnology, specializing in the development of proteomics tools.

Others were relatively slow to follow. The Danish Ministry of Research promoted several programmes in the early and mid-1990s, and Australia's Ministry of Science and Technology funded a Proteome Analysis Facility in the mid-1990s. These did much to spark development of proteomic technologies.

But it is only in the past couple of years that funding agencies have generally started to take proteomics seriously as a key post-genomic approach to biological problems.

The UK Biotechnology and Biological Sciences Research Council, for example, has recently funded three centres to establish transcriptomics and proteomics programmes in organisms whose genomes are sequenced, or close to completion. These are the fruitfly *Drosophila*, the yeast *Saccharomyces cerevisiae*, the plant *Arabidopsis* and the bacterium *Streptomyces coelicolor*.

"We see the approach of launching parallel initiatives in transcriptomics and proteomics as one of our most important scientific priorities in the next few years," says Ray Baker, chief executive of the research council.

In France, five new regional 'genopoles' — sites of concentrated genetics research in a variety of animal, bacterial and plant models — are operating a similar formula. Supported by the research ministry, their funding of FF30 million (US$4.7 million) each for three years is divided roughly equally between genomics, transcriptomics and proteomics.

The European Commission is also funding a proteomics programme to follow up its yeast genome sequencing programme.

The US Department of Energy recently launched proteomics programmes on lower organisms considered relevant to energy production (methane gas) or the cause of, or cure for, environmental problems (bioremediation). Targeted organisms, again partially or completely sequenced, include several archaea, bacteria-like organisms that often inhabit extreme environments.

The main US proteomics effort, however, is in the health area, with the most significant programmes being coordinated through the National Cancer Institute (see page 718).

Although Britain's Medical Research Council has not created any proteomics programmes, its chief executive George Radda describes the area as "important and exciting", and adds that the council "expects to get lots of applications for proteomics projects through our normal mechanisms".

Germany has not thrown itself into proteomics to a great extent, although ironically 2D gel technology was invented in 1975 by Joachim Klose, now at the Max Planck Institute for Molecular Genetics in Berlin. Last year the federal research ministry awarded DM14 million (US$7.3 million) for five years to set up a proteomics centre in Rostock to improve techniques and evaluate them in clinical projects in east Germany.

But this is a fairly isolated action. No dedicated proteomics project will be funded in the second round of the federal government funded German Human Genome Programme, even though this has prioritized post-genomic technologies.

Peter Jungblut, protein-analysis group leader at the new Max Planck Institute for Infection Biology in Berlin, uses proteomics approaches to define new targets for vaccine development. He says the attitude in Germany is "disappointing — but it is typical that Germany waits for new ideas to be developed in other countries before catching on".

Japan, like Germany, is hesitating before embracing large-scale proteomics. The Science and Technology Agency is requesting a ¥284 million (US$2.7 million) investment in proteomics research from the government.

"But the government is reluctant to make a full commitment to proteomics," says Teruhisa Noguchi, former director of the Helix Research Institute, a genomics centre that is contemplating a shift in emphasis towards proteomics, whether or not major government investment comes through. Helix is funded by the Ministry of International Trade and Industry and ten pharmaceutical companies.

Switzerland has no funding specifically earmarked for proteomics. But the Geneva-based Swiss Institute for Bioinformatics is a proteomics flagship for the country, looking after several public protein databases. Headed by Denis Hochstrasser, the institute is financed by the city of Geneva, and is run on proceeds from a new commercial arm, GeneBio, which licenses pharmaceutical companies access to its protein databases (see *Nature* **394,** 214; 1998). GeneBio returns 75 per cent of its profits to the institute.

MRC/STRAT MASTORIS

Radda: expects many grant applications.

The pharmaceutical industry is watching developments with interest. Most companies have modest proteomics programmes in-house, and often collaborate with academic groups on projects likely to drive the technology. Many use service companies, such as Oxford GlycoSciences and Large Scale Biology, for larger clinical projects. But, before investing in large in-house programmes, they are waiting to see if proteomics fulfils its potential as a discovery tool. ■

© 1999 Macmillan Magazines Ltd