

# PROTEIN ARRAYS

## new technology for the proteomics era

The study of the human proteome will drive drug discovery in coming years; although how, when, and in what direction are a little uncertain at present. To build understanding of the role of the human protein complement in health, development, and disease, a database more complex than the human genetic sequence is under construction. To face the challenge, new technologies for protein analysis must be developed that are faster and provide more information than current approaches.

**By Dr Jeffrey P. Kiplinger**

Two years ago, including the word 'genomic' in a corporate press release was the magic recipe for funding a biotech company. Today, the words 'proteome' and 'protein' have nearly the same power and effect. Plentiful news articles offer exciting predictions for the future of drug discovery in the Era of Proteomics; reminiscent of the days barely five years ago when researchers began to see that the end was in sight for the completion of the human genome sequence. It is easy to forget the challenges that remain for those who wish to capitalise on the knowledge that will come from the completed genome.

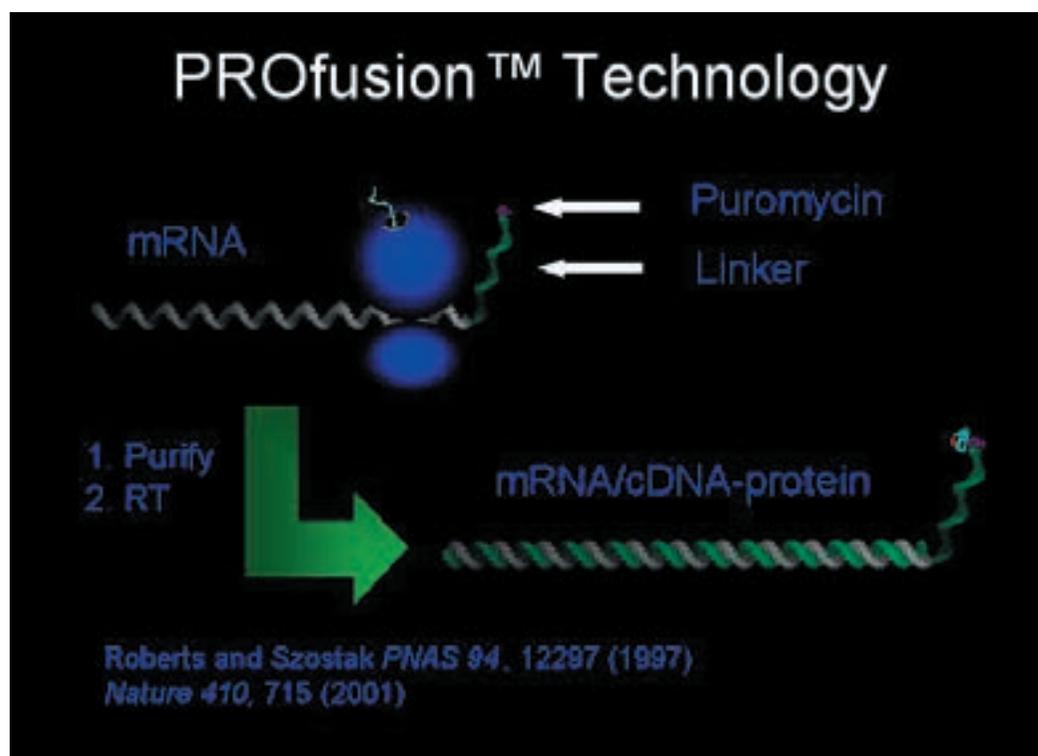
With a human complement of only 35,000-40,000 genes now predicted, and the diversity of the proteome clearly much larger; genetic sequence matching appears to be only one of many tools that will be used to understand protein expression.

The drug discovery process begins with validated targets. Validation is still a process bottleneck; requiring understanding of protein function and protein interactions that are not easily elucidated.

Development of tools for rapid analysis of protein mixtures is proceeding at a frenetic pace, but there are still technological hurdles. The dynamic range and quantitative ability of most methods are less than researchers need to gather all the data they would like.

The ability of the industry to address all three of these issues is growing rapidly, but at present the foundation of our developing understanding of proteome diversity and protein function is in protein analysis. When the word 'proteomics' is tossed around in the press, it generally connotes the analysis of protein mixtures from tissues, both specific to a disease/condition and controlled. The objective is to rapidly identify new or previously known proteins associated with the condition, and to understand how their level of expression and interactions with other proteins are important.

Two-dimensional gel electrophoresis, developed over the last 25 years, is the core technology for proteomic analysis. The value of 2D-gel methods has increased recently due to improvements in the gels themselves that increase resolution and reproducibility<sup>1</sup>. Integrated technologies, such as software that facilitates gel imaging and evaluation, and automated mass spectrometric evaluation of purified proteins resolved on gel plates, have increased the information throughput of proteomics<sup>2</sup>. However, gel electrophoresis remains limited by the labour-intensive sample preparation and gel handling processes, and by the inability to work with certain proteins or to detect those in low abundance.

**Figure 1**

Phylos' PROfusion™ technology covalently links proteins to their mRNA coding sequences. Affinity purification of active proteins is followed by reverse transcription to cDNA, which can be amplified and used to express active proteins or epitopes

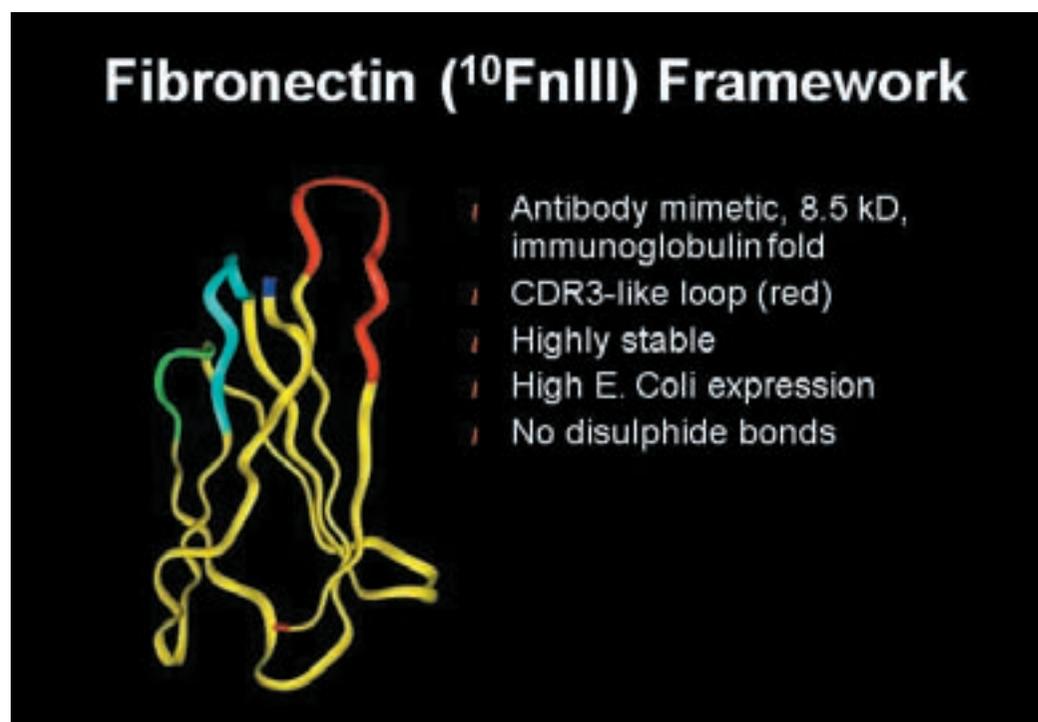
The rapid development of microarray technologies for genomic research, particularly in the analysis of differential gene expression, has piqued the interest of a number of proteomics researchers. Expression profiling – measuring the location, timing, causative factors, and level of protein expression – is everything to proteomics researchers. Because the number of human proteins is expected to be several times greater than the number of genes, and because the same protein may have different functions in different environments or at different times, the information set needed to map out an understanding of the proteome may be several orders of magnitude greater than the genome.

The promise of microarrays lies in the spatially addressable grid of specific binding sites, implying that hundreds or thousands of unique binding events might be analysed simultaneously. Protein microarrays might be used to examine many protein-protein, protein-ligand, or enzyme-substrate interactions on a single slide or 'biochip'. While the difficulties inherent in protein production and purification will limit the scale of protein arrays compared to DNA, single point resolution of specific binding sites will still provide real value. This promise has several young companies excited about the possibility of supplementing, or possibly replacing, the current 2D-gel technology with 'protein chips'.

### Challenges for chip makers

The most easily understood application for protein arrays is in expression profiling – the measurement of the variation in expression of known proteins within tissues or cells; over time; or in response to challenge by drugs, toxins, injury or disease. It has proven difficult to predict protein expression from mRNA expression<sup>3</sup>, limiting some early enthusiasm for using DNA arrays in understanding the proteome. As knowledge of the proteome grows, a database of specific marker proteins and post-translational modifications to marker proteins should allow construction of investigative and diagnostic arrays that are designed to bind these protein targets.

Although the manufacture and use of DNA arrays has been automated and standardised, there are some important hurdles that still exist for developers of protein arrays. Proteins are not easy to attach to surfaces, at least not if the hope is to offer a consistent density and orientation of binding sites for ligands. Some proteins are easily denatured at solid-liquid and air-liquid interfaces, rendering protein arrays much less stable than DNA arrays. Inconsistent behaviour of sites on an array renders the array useless for quantitation, and decreases the selectivity and specificity of the sites. Without highly specific binding to proteins on the chip, thorough washing of the array may remove

**Figure 2**

Phylos' protein microarrays incorporate selected proteins into antibody mimetics based on a fibronectin framework. The highly stable fibronectin allows the production of dry, heat-stable arrays for easy shipping and use

specific binders as well as non-specific binders. Non-specific binding of proteins can be a problem at micromolar target concentrations. To obtain good information content in a protein probe array requires that the presented epitopes bind specific targets at nanomolar concentrations.

Quantitative protein expression profiling is arguably much more important in developing a complete understanding of the proteome than in gene expression profiling. While the quantitative multicolour fluorescent and radiographic detection schemes developed for genomic arrays are being adapted by protein array developers, their productive use and information content is limited by the aforementioned difficulties with obtaining consistent binding results across the array.

Expression profiling and diagnostic arrays are being developed by several companies, including LumiCyte (Fremont, CA), Zyomyx (Hayward, CA), Axcell BioSciences (Newtown, PA), BioSite (San Diego, CA) in collaboration with Large Scale Biology Corporation (Vacaville, CA), and Phylos (Lexington, MA). At a first look, these competitors intend to market commercial protein chips, frequently by binding antibodies to known proteins to arrays. Proprietary technology approaches differ, but perhaps the most important distinction among this group is the database of protein structures that can be used as antigens to produce the antibody arrays. Zyomyx also leverages licences to

phage display libraries from Dyax (Cambridge, MA) and Cambridge Antibody Technology (Melbourn, UK); libraries they will use to expand their available selection of arrays.

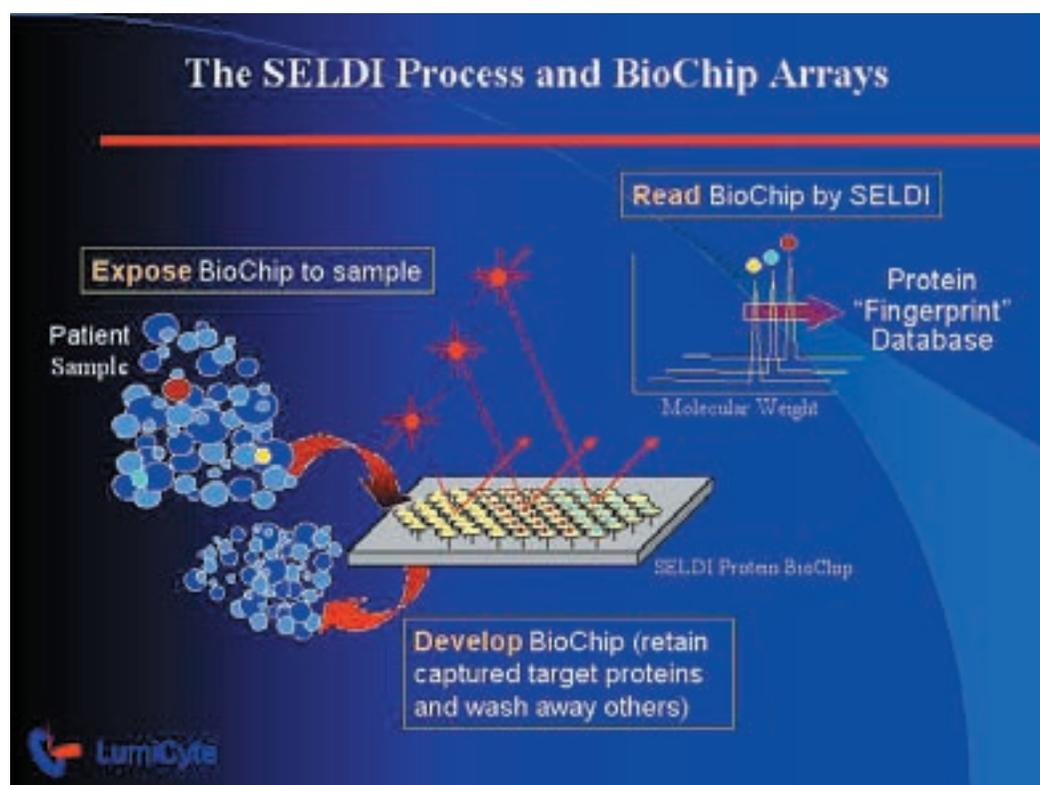
Phylos has taken a combinatorial approach to generating displayed binding epitopes, using patented PROFusion™ technology to produce libraries of billions of small proteins fused to their coding mRNA sequences (Figure 1). Affinity selection among these large libraries is used to obtain a set of useful probes, and the attached mRNA can then be used to generate DNA to express the proteins in standard cellular vectors. Perhaps the most interesting feature of Phylos' approach to chip construction is to express the binding domain as part of a fibronectin isoform that Phylos selected to mimic an antibody (Figure 2). Phylos has shown that this antibody mimic is not only highly selective for the target antigen but is highly heat stable, producing arrays that can be stored and shipped dry without loss of binding activity.

Founded in 1997 with exclusive rights to technology developed at Massachusetts General Hospital, the company now boasts more than 80 employees. Richard Wagner, senior vice-president of Research at Phylos, says that the company's HIP chip™ should be available this year to collaborators. "The chips can be produced to bind, for example, a set of cytokines associated with a clinical condition, and a large pharmaceutical

## Proteomics

**Figure 3**

LumiCyte's microarrays use chemical surface modifications, rather than surface-bound proteins, to capture proteins from biological samples. The arrays allow direct ionisation of the captured proteins in a mass spectrometer, generating a data set that 'fingerprints' the protein content of the sample



company can profile a large set of clinical samples." But Wagner notes that the company's PROfusion™ technology can be used to produce libraries of proteins from genomic DNA as well as through combinatorial processes. "This opens up target validation as a potential application. Although we don't plan at present to do this on chips, some of the ideas being developed in other research groups are encouraging us to think about other ways of displaying protein libraries. For now, we've chosen the fibronectin antibody mimic because it allows us to produce a stable commercial product."

### New technology and old

Proteome Systems (Sydney, Australia and Woburn, MA) has discovered a way to effectively build microarrays out of spots on 2D gels. In an effort to better automate the 'traditional' approach to protein profiling, the company developed its own 'gel chips'. These smaller (the size of a 96-well plate) versions of standard format gels may be processed faster and achieve similar resolving power. Proteome Systems further improves gel chip performance by pre-fractionating proteins according to isoelectric point, and by removing high-abundance interferences such as serum albumen. But the company's unique contribution is not

the gel chip itself; it lies in what happens to the gel chip after development.

Malcolm Pluskal, executive vice-president of new technology and business development for Proteome Systems, explains that: "Each gel chip is developed and stained, and the spots are blotted using standard techniques on to a polymer sheet. We've developed a device called the Chemical Printer™, in collaboration with Microfab (Austin TX), that prints an array of chemical or biochemical reagents right on to each spot of interest on the blot. We can then use a variety of detection methods, including adding MALDI matrix and ionising the spots directly in a mass spectrometer." So far the Chemical Printer has been used with enzymes for digestion of the blotted protein. The ability to print up to 100 different reagents in a grid on to a protein spot allows several different digestions to be performed, and the amounts of enzyme applied can be varied to help minimise enzyme artifact peaks in the mass spectrum. "We can even use the mass spectrometer to spatially resolve incompletely separated proteins within a single spot," says Pluskal.

Proteome Systems intends to develop an integrated technology platform for protein discovery work called ProteomIQ™, expected to launch in the third quarter of 2001. The platform will incorporate gel

## Proteomics

### References

- 1 Gorg, A et al. The Current State of Two-Dimensional Electrophoresis with Immobilized pH Gradients. *Electrophoresis* 2000;21: 1037-53.
- 2 Wilson, JF. New Technology Spurs on Proteomics. *The Scientist* 2001;15(7):12.
- 3 Anderson, NL, Anderson, NG. Proteome and proteomics: new technologies, new concepts, and new words. *Electrophoresis* 1998;19: 1862-71.

chips, the Chemical Printer, reagents from Sigma-Aldrich, and Shimadzu's AXIMA-CFR MALDI mass spectrometer. "The chemical printer can apply anything to a protein spot – chemical or biochemical reagents, modifiers, small molecule binders – giving us access to lots of applications. We call the 2D gel a non-determined array, but on each spot we can apply a determined array of chemicals to do something different to the protein at each location."

### An alternative to gels

Ciphergen (Fremont, CA) sells instruments and consumables (biochips) for use in biology research laboratories. In 1997 Ciphergen went to market with a 'protein chip', a small array with different surface modifications at each location. Proteins bind to the surface and are ionised directly with a mass spectrometer using a technique called SELDI (Surface Enhanced Laser Desorption Ionisation). Ciphergen has developed a successful integrated system for protein analysis using this technology, and completed its initial public offering last year, raising \$150 million.

Both Ciphergen and LumiCyte (also in Fremont, CA) derive their respective rights to the SELDI process from Molecular Analytical Systems (MAS), a company founded in 1993 by T. William Hutchens. LumiCyte is expanding SELDI applications to build a massive protein discovery database for the purpose of supplying aggregated and integrated protein biology knowledge to its collaborators. LumiCyte leverages strategic partnerships with clinical organisation and pharmaceutical companies to profile proteins from, ultimately, hundreds of thousands of clinical samples. At the heart of LumiCyte's technology is a new type of biochip that presents surface binding sites for proteins in a dense array, allowing LumiCyte to develop a rapid system for protein profiling (Figure 3). The binding sites are chemical modifications, not antibodies, that provide stable and reproducible affinity arrays that are highly target-specific.

Anthony D. Bashall, LumiCyte's vice-president of business development, says that LumiCyte will build its database initially for partners who will use the knowledge for drug development. Later the company will move beyond the laboratory to point-of-care diagnostic applications. "Some of our partners have described us as the first 'post-proteomics' company, in that what we're building will be immediately useful in the delivery of more personalised medicine. We're already using this technology and information for early detection of prostate cancer and diabetes in asymptomatic indi-

viduals, and to monitor individual response to therapy." LumiCyte and Kratos Analytical (Manchester, UK, a division of Shimadzu Biotech) recently announced a partnership that will result in the installation of 105 high-throughput laser desorption mass spectrometers in LumiCyte facilities. With this technology investment, LumiCyte has announced its intention and ability to profile disease markers and phenotypic indicators across many thousands of individuals. The approach is similar to what pharmacogenomics companies are attempting in building phenotypic gene expression profiles.

"It is a completely orthogonal approach to 2D gel profiling, and has tremendous potential because it is so fast and scalable," notes Bashall. "We will be able to use our technology base to look at hundreds of clinical samples a day. We have this technology edge, but we're going to use it to go after the real value – new medical knowledge."

A shared sentiment among all these young companies is that the field of Proteomics offers plenty of room for players. The expectation is that a complete description of the proteome will require a database several orders of magnitude greater in size and complexity than the genome, and that multiple approaches will be required to build that database. The proteome is not a sequence, but a set of interdependent variables that will take years to fully understand. "The real challenge coming up is data management," says Pluskal. "Almost everyone is working to create new tools to help with understanding the columns and columns of data. The way the data is presented back to the scientist is important to its understanding. It's a human ergonomics challenge." DDW

---

*Dr Kiplinger received his PhD from Indiana University, USA. He spent 10 years with Pfizer Pharmaceuticals, where he developed open access LCMS and automated LCMS purification technologies during the 'combi-chem revolution'. In 1998 he founded the Gilson Inc Centre for Integrated Discovery Technology, acting as its Scientific Director. He is currently president and founder of Pragmatic Approach which helps scientists develop new business opportunities.*