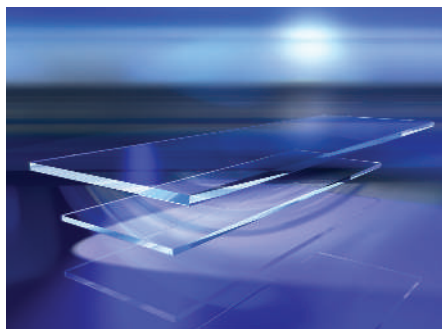


Growing pains

Protein microarrays are coming of age, and the development of specialized technologies is extending their high-throughput capabilities. Michael Eisenstein reports.

Like the younger child who has had an older sibling to 'soften up' his or her parents and make life a little easier, protein microarrays have benefited from lessons learned during the noisy adolescence of DNA microarrays. "Intellectually, the assays are identical to DNA microarrays," says microarray pioneer Mark Schena, visiting scholar at TeleChem International in Sunnyvale, California. "The basic tenets of miniaturization, automation and parallelism all hold for DNA and protein." Indeed, some tools of the trade, such as array spotters and readers, have needed virtually no changes to make the move to proteins. But the differences are there nonetheless, and the protein-microarray field is rapidly developing its own identity, with new challenges that require a diverse set of specialized tools and tricks.

One key distinction emerges early in the analysis process: the selection of array substrate. Unlike oligonucleotides, proteins are broadly heterogeneous in size, shape and chemistry, and the diversity of applications for protein arrays means that users need to shop around for the appropriate substrate. "We usually test about three or four surfaces, and see which one



Schott's Slide H is coated with a hydrogel that is a three-dimensional binding substrate for proteins.

gives us the best signal-to-noise ratio for a given assay," says Michael Snyder of Yale University.

There has been an explosion of options in the market, and several vendors — such as Xenopore of Hawthorne, New Jersey, and Schott of Jena, Germany — specialize almost exclusively in slides and substrates. Functionalized glass surfaces are a strong choice, as they virtually eliminate background fluorescence. According to Schena, TeleChem's SuperEpoxy glass slides are among the company's most popular prod-

ucts for protein-array construction. "The key benefit of the epoxide is that it's highly reactive at physiological conditions," he says. "Just depositing proteins on an epoxide-coated surface results in covalent linkage to the glass."

Three-dimensional substrates are also popular. For example, the Nexterion Slide H from Schott is coated with a covalently linked three-dimensional hydrogel with reactive groups that readily bind to proteins or peptides. "Slide H preserves the three-dimensional structure of proteins, providing a hydrophilic cell-like or cytosol-like environment, thereby maintaining stability and functionality," says Rüdiger Dietrich, Schott's director of research and development and technical support. PamGene of 's-Hertogenbosch in the Netherlands also takes advantage of a three-dimensional environment for its flow-through array platform. In its set-up, samples are pumped back and forth through a porous inorganic substrate where capture probes are immobilized.

Nitrocellulose, a timeless classic for protein work, remains among the most popular substrates. "We still don't have a substratum that's better than nitrocellulose for its binding

LOSING THE LABEL

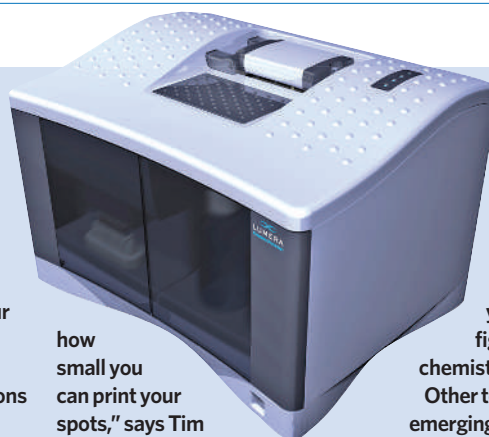
The appeal of using arrays free from fancy adornments — fluorescent, radioactive or otherwise — is fairly obvious. Such arrays could eliminate extra work and reduce errors in detection and analysis.

Surface plasmon resonance (SPR) is a promising technology for this approach. In SPR the interactants are fixed to a gold-coated substrate, and sample binding is detected as mass concentration-dependent changes in refractive index at that spot, which makes it possible to monitor binding in real time. "You can look at specificity, affinity, kinetics, make concentration measurements, and work in a range of different sample environments," says Gary Franklin, industrial-sector specialist at Biacore.

Recently acquired by GE Healthcare in Little Chalfont, UK, Biacore has pioneered the development of SPR platforms for a wide variety of proteomics applications. It has two main

options for array users. In the Flexchip platform, 400 interactants can be spotted on a slide and then screened against a single sample. In contrast, the A100 is limited to 20 immobilized interactants but, thanks to parallel flow systems, it can perform up to four simultaneous screens with large numbers of samples. "The A100 can look at up to 3,800 interactions per day in a variety of different, multiplexed ways," says Franklin.

Lumera of Bothell, Washington, is a relative newcomer to the market. It takes advantage of extremely rapid optical-switching technology, originally developed for telecommunications, in its ProteomicProcessor SPR instrument. Throughput was a key limitation of early SPR array experiments, and Lumera says its switching technology has reduced the problem. "We're basically limited by the size of the slide and



The ProteomicProcessor from Lumera will offer high-throughput label-free analysis.

how small you can print your spots," says Tim Londergan, director of the company's bioscience business unit. "You can really see this tracking to DNA microarray scale." Lumera has nearly finished testing its 'beta' instrument and plans to launch its first commercial system in January 2007.

But SPR is not without limitations, and reduced sensitivity remains a common complaint. "My feeling is that there may be a somewhat limited dynamic range," says Joshua LaBaer, director of the

Harvard Institute of Proteomics. "It's a fairly narrow window, and you have to be able to figure out how to get your chemistry within that window."

Other technologies are also emerging, such as the arrays based on microcantilevers made by Protiveris of Rockville, Maryland, or the atomic-force microscopy system from BioForce Nanosciences of Ames, Iowa. Most of these platforms are in their infancy, but technology development is under way and many microarray users foresee a label-free future — one way or another. "It's going to revolutionize the way we think about things," says LaBaer. "It won't happen in ten days or a year, but it's going to happen." M.E.

capacity — it's cheap and can be mass-produced," says Emanuel Petricoin, co-director with Lance Liotta of the Center for Applied Proteomics and Molecular Medicine at George Mason University in Manassas, Virginia. "The problem is that it has high background fluorescence and you have to come up with different labelling strategies to get around that." One solution uses an ultrathin nitrocellulose layer — such as the PATH slides from GenTel BioSciences of Madison, Wisconsin — which maintains binding capacity but reduces background noise.

More specialized surfaces enable orientation-specific presentation of protein probes, such as nickel-NTA (nitrilotriacetic acid) surfaces for use with polyhistidine-tagged proteins. Lumera in Bothell, Washington, uses a protein-protein interaction-based immobilization format for its arrays. "We have a proprietary protein-tag technology that's composed of two coiled coils that bind together with picomolar affinity," says product group manager Ronald Dudek. "One can be engineered into a protein or antibody library, the other is part of the surface chemistry." The pairing of streptavidin-coated slides with biotinylated proteins is also widely used.

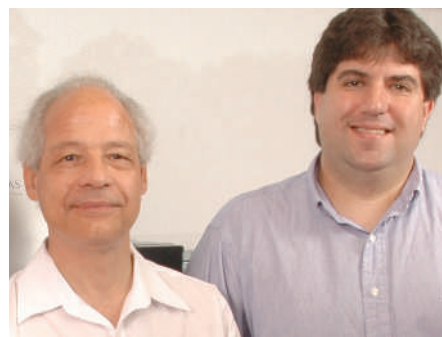
Seeing the light

One aspect of arrays that has changed relatively little is the dominance of fluorescence as a method of detection. "People generally use the standard Cy3 and Cy5 fluorophores, which are cheap, reproducible, and scanners are already configured to read them," says Petricoin. His group has done initial studies with

semiconductor quantum dots, which offer the benefits of robust long-term fluorescence and enhanced potential for multiplexing. But for many investigators, generically reactive fluorescent dyes are sufficient. "In some ways, labelling is easier than with DNA," says Schena, "because proteins are naturally highly reactive with fluorescent reagents."

Some manufacturers improve the quality of fluorescent-array experiments with specialized instrumentation. The ZeptoREADER from Zeptosens in Witterswil, Switzerland, uses 'planar waveguide' technology, in which an evanescent electromagnetic field is generated by directing light into a specially designed array substrate through a diffraction grating. "The key benefit is that you excite only fluorophores that are bound on the surface," says managing director Markus Ehrt. "In many conventional assays, you don't need to separate bound from unbound fluorophores." Alternatively, in PamGene's flow-through system, samples are repeatedly cycled through porous arrays while being imaged by a sensitive charge-coupled device camera, making it possible to take real-time kinetic measurements.

For experiments involving particularly small samples or scarce protein targets, it might be necessary to amplify the signal. One way to do this is rolling-circle amplification (RCA), first adapted for use with protein arrays by David Ward at Yale University. In RCA, circular DNA molecules are hybridized to capture probes that are conjugated to a detection antibody. These are then repeatedly replicated by DNA



Lance Liotta (left) and Emanuel Petricoin use reverse-phase arrays instead of sandwich arrays in their clinical research.

polymerase and detected with complementary fluorescent probes. Another alternative involves enhanced chemiluminescent detection with tyramide signal amplification, a system available from PerkinElmer of Wellesley, Massachusetts.

As effective as these techniques have proven, they all require a degree of tagging or modification, which can be time-consuming or even impractical for certain samples, such as clinically derived preparations. So many researchers are keeping their eyes on the evolution of 'label-free' technologies that could greatly simplify future studies (see 'Losing the label').

As with genomic arrays, the earliest application — and still the main use — of protein arrays has been the detection or quantification of targets from a research sample or clinical preparation. This typically involves a 'capture array'

AN APT SOLUTION?

Larry Gold has a solution for the problems faced by users of antibody arrays — stop using antibodies. Gold, founder and chief executive of SomaLogic in Boulder, Colorado, is outspoken in his belief that aptamers — small nucleic acid molecules with specialized functional properties — should replace antibodies in capture arrays.

Years ago, Gold and his team tried to calculate how much multiplexing could be done with an antibody array before background noise became a significant issue. "We guessed that sandwiched antibody arrays were going to start running into serious noise problems when there were 20 analytes per array," Gold says. He thinks that aptamers could provide a level of specificity that surpasses what most antibodies achieve.

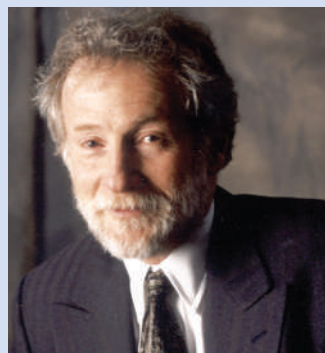
More than 15 years ago, Gold was one of the inventors of a now widely used procedure called SELEX — systematic evolution of

ligands by exponential enrichment. In this system, multiple rounds of selection and amplification can be used to select for DNA or RNA molecules with high specificity for a target of choice.

Gold and his colleagues have since enhanced the procedure, and SomaLogic uses a high-throughput version of SELEX to generate 'photoaptamers', which can be covalently crosslinked to bound targets following irradiation. As a result, aptamers with high affinity and specific crosslinking can be used to measure proteins in complex samples without needing the extra specificity provided by secondary antibodies.

Another advantage of the aptamer platform is simplicity of detection — once protein targets are bound, they can be labelled with a generic protein-binding fluorescent dye, eliminating the need for analyte-specific sandwich detection reagents.

Early platform tests have been promising — SomaLogic is achieving success rates of about 80%, Gold says, and he believes a product launch is imminent. "With the right partner, we could launch a product for the research market within some months," he says, "but we think diagnostics is key, and we're even more interested in transforming evidence-based healthcare."



Larry Gold believes his aptamer arrays will succeed in applications where antibody arrays fall short.

Some antibody experts are impressed by aptamers, but doubt whether they will unseat the current king. "What Larry Gold and others have shown is that when it works, it works very, very well," says Mathias Uhlén of the Royal Institute of Technology in Stockholm, Sweden, "although I would be surprised if aptamers are the dominant scaffold in the future, due to limitations in the chemical space."

But SomaLogic is banking on the use of novel nucleotides to increase this chemical space, and believes that a full proteome complement of aptamers is unnecessary — so that high specificity against a few thousand well-chosen targets could be more than sufficient. "We think that, for diagnostics, there's so much redundancy in biology that you'll be able to do useful biomarker discovery with an incomplete proteome that's still quite large," says Gold.

M.E.

in which selected antibodies are spotted on to a substrate and are then exposed to the sample. Detection can be performed directly with a labelled sample, although this can increase the background noise. Many researchers prefer 'sandwich' formats in which a second, tagged antibody is applied that recognizes a different epitope from the capture antibody. This considerably improves sensitivity and the signal-to-noise ratio, but also requires two high-quality antibodies for every target.

Antibody arrays have been commercially available for some time, but companies have taken different approaches to them. Sigma-Aldrich of St Louis, Missouri, for example, has several arrays in its Panorama product family, each containing 100–200 antibodies targeting proteins involved in processes such as signal transduction or gene regulation. The company also plans to release a broad-content array with more than 700 antibodies. "Our current pathway arrays will be included, in addition to various other signalling and regulatory gene-product antibodies," says market-segment manager Richard Pembrey.

TeleChem, on the other hand, leaves antibody selection to the user for its ArrayIt microarrays. "For now, it's custom work just because a full proteomic complement of antibodies is still expensive for most researchers," says Schena. "But our printing technology scales pretty nicely, and our products can span a pretty wide spectrum, between 100 elements and 50,000."

After encountering limited success with capture arrays in their clinical research, Liotta

and Petricoin developed the reverse-phase protein array, in which samples are spotted on the array and then probed with detection antibodies, requiring only one antibody per analyte. Much of their work has centred on signalling pathways in human disease, and this technique has worked well with their research, they say. "From a few thousand cells obtained by laser microdissection, we can look at hundreds of phosphorylation endpoints quantitatively, and look at a target and all the downstream signalling around it," says Liotta.

Such sample-specific arrays are difficult to commercialize, although Zeptosens is attempting to address the needs of this community with its cell lysate array (CeLyA) product line, which provides users with protocols, reagents and equipment to prepare chips for reverse-phase experiments. The company also has an active service division. "We deposit roughly the content of one or two cells per spot," says Ehrat, "and we can monitor changes of 15% from control to a treated sample."

But both approaches face a key limitation: antibodies. "A lot of users have two problems," says Mathias Uhlén, a researcher at the Royal Institute of Technology in Stockholm, Sweden, and chair of the Human Antibody Initiative of the Human Proteome Organisation (HUPO). "One is that you buy reagents, and half of them don't work in your application, and the second is that it's not easy to buy 200 antibodies that work on a single platform." Part of the solution lies in thorough validation, but other problems arise from the feature density of today's arrays

and the broad range of protein expression. "Even if you have an antibody with picomolar affinity for your target, where the background is micromolar; if you have cross-reactivity against a protein that is 10^6 times more abundant, you will see that protein first," Uhlén says. Some are exploring alternative affinity reagents, such as recombinant single-chain antibodies or nucleic-acid aptamers (see 'An apt solution?'), but most in the field still see the limits of antibodies as secondary to their strengths. "So far, the good old antibody is still going strong," says Uhlén.

All together now

High-content protein microarrays have brought the classic protein-protein interaction assay to levels of throughput previously only possible with two-hybrid assays. These 'proteome chips' originated in Snyder's lab, in the form of arrays composed of protein products from nearly 6,000 yeast open reading frames. These chips are now available from Invitrogen of Carlsbad, California, which has continued to develop these and other arrays as part of its ProtoArray product line. Invitrogen sells both human and yeast proteomic arrays in a variety of formats, designed for use in protein-protein interaction assays as well as functional studies.

Developing arrays of soluble — and ideally, functional — protein at this scale poses considerable challenges, says Paul Predki, vice-president of proteomics at Invitrogen. "Our latest array product has more than 8,000 human proteins, and you can imagine the challenges

(ALMOST) NO ASSEMBLY REQUIRED

Each stage in the building of a protein chip — expression, purification, immobilization — adds a layer of experimental complexity, as each feature may need its own optimization process to ensure consistent quality. "We thought we needed a better way to do this," says Joshua LaBaer, director of the Harvard Institute of Proteomics.

The solution that he and his team arrived at was the 'nucleic acid programmable protein array' (NAPPA), in which cDNAs encoding GST fusion proteins are arrayed on chips alongside antibodies that recognize GST. The array is then subjected to cell-free transcription and translation; as protein is produced, it gets bound by an antibody and presented for analysis. According to LaBaer, NAPPA has simplified his group's research. "You don't have to purify proteins — you just purify DNA, so it's pretty easy, and it's been successful for printing about 95%



Joshua LaBaer uses self-assembling arrays to bypass problems with proteins.

to 96% of the things we make," he says. "And when we do protein-protein interactions, we're getting interactions that make sense and not a lot of false positives."

Although initial arrays were limited in size, LaBaer and his team have since generated NAPPA arrays with up to 2,000 features, and they hope to surpass this soon.

Other techniques even bypass DNA immobilization. In the protein

in situ array (PISA) developed by Michael Taussig of the Babraham Institute in Cambridge, UK, cDNAs are amplified *in situ* with primers that encode polyhistidine tags, so that proteins can be captured on a nickel-NTA-coated surface.

More recently, Philipp Angenendt of the German Cancer Research Center in Heidelberg transferred the PISA principle to a microarray set-up, integrating cell-free

production of histidine-tagged protein from unpurified PCR fragments with a multiple spotting technique (MIST) previously developed by his group. MIST uses automation to apply array reagents precisely and sequentially to specific spots. As a result, each transcription/translation reaction is confined to a tiny, sub-nanolitre droplet, allowing greater density — up to 13,000 spots at present. "The nice thing about it is that the proteins expressed remain in a liquid environment," says Angenendt, "and the structure should be as intact as it can be with a solid-phase immunoassay format."

Both Angenendt and LaBaer are now fine-tuning their processes. "We're really working on large-scale screens, doing biomarker and protein interaction studies," says LaBaer. "We've also got some preliminary enzymatic data that look promising."

M.E.



JPT uses a high-throughput, non-contact printing process to assemble its peptide arrays.

of doing quality control for the functionality of all of these proteins, especially when for some 15% we don't even know the function, and other proteins have multiple functions," he says. Invitrogen bypasses some of the risks of misfolding or improper processing by expressing proteins in insect cells rather than in bacteria. "Almost all of our human proteins are expressed using baculovirus and we've found, for example, much higher success rates in obtaining active kinases," says Predki.

Snyder is also continuing his work with the yeast arrays, which now cover roughly 75–80% of the yeast proteome. These proteins have been cloned as both C- and N-terminal glutathione S-transferase (GST) fusions to improve the odds of obtaining well-folded fusions, although some proteins refuse to comply. Several groups are now attempting to resolve this issue with chip formats in which transcription and translation of cloned genes are performed on the chip (see 'Almost no assembly required').

Proteome chips also have clinical promise, and Snyder suggests that they could be valuable for drug screening. "Any drug company with a set of lead compounds should put them on a protein chip and see what the possible side effects might be, or what their targets are," he says. "I think that will be the future."

Pathogen-derived proteome chips may also have clinical value — Snyder recently developed a coronavirus protein array for diagnostic use, and Joshua LaBaer, director of the Harvard Institute of Proteomics in Boston, Massachusetts, has developed full proteomic arrays for the pathogens *Francisella tularensis* and *Vibrio cholerae*. "The idea would be to take serum from individuals who have been infected, both pre- and post-immune response, and look at

the proteins against which the patient has mounted a response," says LaBaer.

JPT Peptide Technologies in Berlin, Germany, is examining similar applications, with microarrays displaying overlapping peptides comprising the full *Mycobacterium tuberculosis* proteome. "We started with a small subset of 10,000 peptides from tuberculosis markers," says managing director Michael Schutkowski, "and found that the chip is more predictive and sensitive than all the other tuberculosis tests on the market."

Where form meets function

Functional information is available for only a subset of the more than 23,000 proteins in the human proteome, and much of this is incomplete or even anecdotal. A truly comprehensive understanding of biological processes requires the building of functional proteomic maps, and protein arrays could be a perfect tool for this.

Snyder and his team made a step forward when they applied their proteome chips to the assembly of a yeast phosphorylation map, incubating their microarrays with 87 different yeast protein kinases, and using radiolabelled ATP to identify proteins that were being modified. "We showed that closely related kinases had different specificities," says Snyder. His lab was also able to combine its findings with other genomic and proteomic data to identify regulatory 'modules' of interacting proteins that seem to be conserved in other eukaryotic species. "In my mind, that's the power of '-omics,'" he says. "We can come up with new principles by looking at these large data sets."

Invitrogen's arrays are also suitable for such assays. "The same application is fully validated on our human array," says Predki. And Sigma-Aldrich has a product for kinomic analysis, the Panorama Human Kinase v1 array, based on technology licensed from Procognia in Maidenhead, Berkshire, UK. This array incorporates 152 different human kinases for the identification of interacting partners and substrates, or the analysis of putative therapeutic compounds. On the clinical side, Liotta and Petricoin have found success using phosphospecific antibodies with reverse-phase arrays to characterize differential phosphorylation in biopsy specimens, and their arrays are now undergoing clinical trials as a diagnostic or prognostic tool for cancer.

Several groups are also developing array formats for glycan analysis. Brian Haab of the Van Andel Institute in Grand Rapids, Michigan, has used lectins — a family of proteins with specific binding preferences for different types of glycans — as a probe for monitoring glycosylation patterns of array-bound proteins. "We now work with about 20 different lectins," says Haab, "and we

recently demonstrated that we can do partial digestion of glycan structures using glycosidases to expose additional underlying structures to lectins, allowing us to get more complete information." These are mainly being used to characterize glycosylation changes in disease states, and Haab has licensed his platform to GenTel for commercial development.

Snyder, meanwhile, is interested in using such arrays for discovery, and has used polyclonal antibodies against yeast glycans to reveal hundreds of previously unidentified glycoproteins. QIAGEN of Venlo, the Netherlands, will soon launch a glycomics array product, the QProteome GlycoArray kit, based on technology licensed from Procognia, which uses immobilized lectins as a capture reagent.

Peptide arrays are a useful alternative for functional assays, and JPT has applied a peptide-synthesis process to develop a variety of array products. "The peptide chip is much more stable compared with a protein chip," says Schutkowski, "and because we have the peptides beforehand in a microtitre plate, we can analyse each peptide for purity and whether it has the right content and the right modifications before we immobilize it." JPT has used literature- and data-mining to design peptide libraries containing thousands of putative kinase targets, which can be used to identify recognition sequences for tyrosine or serine/threonine kinases. The company also offers similar arrays for phosphatase and protease target identification, as well as 'random' arrays that can potentially reveal previously unidentified target sites.

PamGene uses a peptide-based format in its PamChip arrays for analysing the kinetics of tyrosine and serine/threonine kinases against hundreds of arrayed probes. "IC₅₀ results, selectivity data, mechanism of action information and multiple rate constants can all be obtained in one experimental run," says PamGene's vice-president of technology, Rinie van Beuningen.

As these various applications strive towards full maturity, those in the protein microarray field are wrestling with the same issues that have confronted users of DNA arrays — experimental standardization and accurate quantification. The former issue is now being investigated by HUPO, which is looking to develop guidelines for experimental design and data annotation. But the key to accurate quantification will most likely lie in technological evolution. "Right now these data are just

semi-quantitative, but I think that with the right technologies they could be made completely quantitative," says Snyder. "I'd like to see a future where people can just think up an experiment, buy an array, and do it."

Michael Eisenstein is the former technology editor for *Nature* and *Nature Methods*.



Michael Snyder is looking to a future when people can think up an experiment, buy an array, and do it.